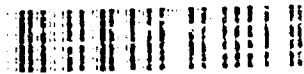
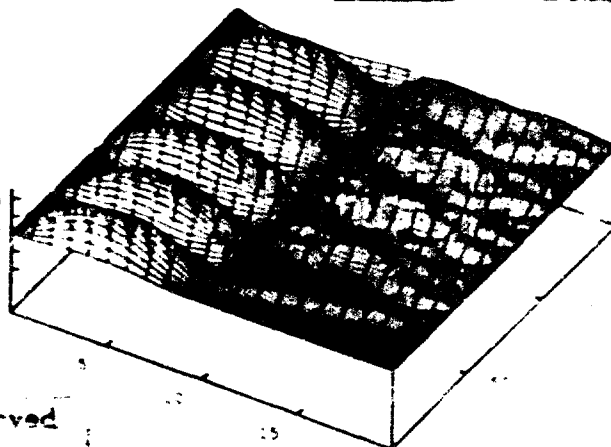
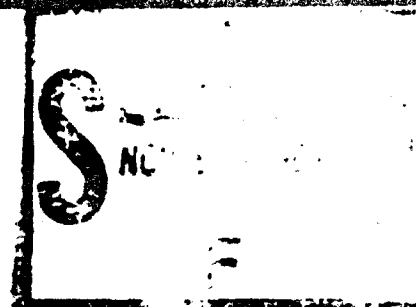


AD-A286 449



**Edited by**

**W. Desch  
F. Kappel  
K. Kunisch**



This document has been approved  
for public release and its  
distribution is unlimited.

**Reproduced From  
Best Available Copy**

**Birkhäuser**

**94-34961**



**94 11 0 1 4 0**

**DTIC QUALITY INSPECTED 8**

**ISNM**  
**International Series of**  
**Numerical Mathematics**  
**Vol. 118**

**Edited by**  
**K.-H. Hoffmann, München**  
**H. D. Mittelmann, Tempe**

# Control and Estimation of Distributed Parameter Systems: Nonlinear Phenomena

International Conference in Vorau (Austria), July 18-24, 1993

Edited by

W. Desch  
F. Kappel  
K. Kunisch



Birkhäuser Verlag  
Basel · Boston · Berlin



Editors

W. Desch and F. Kappel  
Institut für Mathematik  
Universität Graz  
Heinrichstrasse 36  
A-8010 Graz  
Austria

K. Kunisch  
Fachbereich Mathematik  
Technische Universität Berlin  
Strasse des 17. Juni 136  
D-10623 Berlin  
Germany

A CIP catalogue record for this book is available from the Library of Congress, Washington D.C., USA

Deutsche Bibliothek Cataloging-in-Publication Data  
**Control and estimation of distributed parameter systems:**  
nonlinear phenomena ; international conference in Vorau  
(Austria), July 18-24, 1993 / ed. by W. Desch ... - Basel ;  
Boston ; Berlin : Birkhäuser, 1994  
(International series of numerical mathematics ; Vol. 118)  
ISBN 3-7643-5098-9 (Basel ...)  
ISBN 0-8176-5098-9 (Boston)  
NE: Desch, Wolfgang [Hrsg.]; GT

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. For any kind of use the permission of the copyright owner must be obtained.

© 1994 Birkhäuser Verlag, P.O. Box 133, CH-4010 Basel, Switzerland  
Camera-ready copy prepared by the editors  
Printed on acid-free paper produced from chlorine-free pulp  
Cover design: Heinz Hiltbrunner, Basel  
Printed in Germany  
ISBN 3-7643-5098-9  
ISBN 0-8176-5098-9

9 8 7 6 5 4 3 2 1

## Preface

The International Conference on Control and Estimation of Distributed Parameter Systems (Nonlinear Phenomena) was held at the Chorherrenstift Vorau (Austria), July 18 - 24, 1993. The meeting formed a platform for the exchange of new ideas and recent results in control, identification and optimization of nonlinear infinite dimensional systems. It was designed to provide a melting pot for a broad variety of viewpoints, including theoretical aspects like the Maximum Principle, relaxation and stabilizability as well as numerical algorithms for optimization and applications to elastic structures, flow control and population dynamics.

About 45 participants from 9 countries have attended. We thank all of them for their contribution to the success of this meeting. Once again we have enjoyed the pleasant and stimulating atmosphere of the Bildungshaus Chorherrenstift Vorau. Our thanks go to the very helpful staff and in particular to Director P. Riegler who have created an optimal environment for our conference.

We have received funding for this conference from Amt der Steiermärkischen Landesregierung, Graz, Bundesministerium für Wissenschaft und Forschung, Wien, Christian Doppler Laboratorium für Parameter Identifikation und Inverse Probleme, Graz, United States Air Force European Office of Aerospace Research and Development and United States Army Research Development and Standardization Group, London. We greatly appreciate the support from these institutions.

We thank in particular Mrs. G. Krois, whose efforts have been an indispensable help in hundred details of administration work, and who has prepared the TeX manuscript for these proceedings.

Graz, June 1994

W. Desch, F. Kappel, K. Kunisch

## List of Participants

H.T. BANKS	North Carolina State University
V. BARBU	Universitatea "Al.I.Cuza"
J. BLUM	Université Joseph Fourier Grenoble I
J.A. BURNS	Virginia Polytechnic Institute & State University
C.I. BYRNES	Washington University
E. CASAS	Universidad de Cantabria
W. DESCH	Universität Graz
H.O. FATTORINI	University of California
B.G. FITZPATRICK	North Carolina State University
D.S. GILLIAM	Texas Tech University
K.B. HANNSGEN	Virginia Polytechnic Institute & State University
S.W. HANSEN	Iowa State University
M. HEINKENSCHLOSS	Virginia Polytechnic Institute & State University
M.A. HORN	University of Minnesota
L.S. HOU	Simon Fraser University
W. HUYER	Universität Graz
K. ITO	North Carolina State University
F. KAPPEL	Universität Graz
V. KOMORNIK	Université Louis Pasteur
W. KRABS	Technische Hochschule Darmstadt

M. KROLLER	Universität Graz
K. KUNISCH	Technische Universität Berlin
J.E. LAGNESE	Georgetown University
G. LEUGERING	Universität Bayreuth
R. OBER	University of Texas at Dallas
R. PEER	Universität Graz
G. PEICHL	Universität Graz
W. PRAGER	Universität Graz
G. PROPST	Universität Graz
W. RING	Universität Graz
I.G. ROSEN	University of Southern California
T. ROUBÍČEK	Czech Academy of Sciences
E.W. SACHS	Universität Trier
F. SANTOSA	University of Delaware
C. SCHERER	Universität Würzburg
O. SCHERZER	Universität Linz
G. SCHMIDT	McGill University
M.E. TESSITORE	Università di Roma "La Sapienza"
D. TIBA	Romanian Academy
F. TRÖLTZSCH	Technische Universität Chemnitz-Zwickau
C. WANG	University of Southern California
R.L. WHEELER	Virginia Polytechnic Institute & State University

M. YAMAMOTO	University of Tokyo
B.-Y. ZHANG	University of Cincinnati
E. ZUAZUA	Universidad Complutense de Madrid

## Contents

H.T. BANKS, F. KAPPEL AND C. WANG: A semigroup formulation of a nonlinear size-structured distributed rate population model	1
H.T. BANKS AND Y. WANG: Damage detection and characterization in smart material structures	21
M. BERGOUNIOUX, T. MÄNNIKKÖ AND D. TIBA: Optimality conditions for non-qualified parabolic control problems	45
C.I. BYRNES, D.S. GILLIAM AND V.I. SHUBOV: Convergence of trajectories for a controlled viscous Burg- ers' equation	61
P. CANNARSA AND M.E. TESSITORE: Optimality conditions for boundary control problems of parabolic type	79
E. CASAS: Pontryagin's principle for optimal control problems gov- erned by semilinear elliptic equations	97
H.O. FATTORINI: Invariance of the Hamiltonian in control problems for semilinear parabolic distributed parameter systems	115
B.G. FITZPATRICK: Rate distribution modeling for structured heterogeneous populations	131
S.W. HANSEN: A model for a two-layered plate with interfacial slip	143

M. HEINKENSCHLOSS AND E.W. SACHS:	
Numerical solution of a constrained control problem for a phase field model	171
M.A. HORN AND I. LASIECKA:	
Uniform stabilizability of nonlinearly coupled Kirchhoff plate equations	189
K. ITO:	
Boundary temperature control for thermally coupled Navier-Stokes equations	211
J. KAZIMIR AND I.G. ROSEN:	
Adaptive estimation of nonlinear distributed parameter systems	231
V. KOMORNIK:	
Decay estimates for the wave equation with internal damping	253
W. KRABS:	
On the controllability of the rotation of a flexible arm	267
J.E. LAGNESE:	
Modeling and controllability of interconnected elastic membranes	281
G. LEUGERING:	
On feedback controls for dynamic networks of strings and beams and their numerical simulation	301
T. ROUBÍČEK:	
Various relaxations in optimal control of distributed pa- rameter systems	327
F. TRÖLFSCH:	
Convergence of an SQP-method for a class of nonlinear parabolic boundary control problems	343

M. YAMAMOTO:	
Conditional stability in determination of densities of heat sources in a bounded domain	359
 B.-Y. ZHANG:	
Boundary stabilization of the Korteweg-de Vries equation	371
 E. ZUAZUA:	
Controllability of the linear system of thermoelasticity: Dirichlet-Neumann boundary conditions	391



## A SEMIGROUP FORMULATION OF A NONLINEAR SIZE-STRUCTURED DISTRIBUTED RATE POPULATION MODEL

H. T. BANKS, F. KAPPEL, AND C. WANG

Center for Research in Scientific Computation  
North Carolina State University

Institut für Mathematik  
Universität Graz

Department of Mathematics  
University of Southern California

**ABSTRACT.** A variation of the Sinko-Streifer model in population dynamics where besides the observable characteristics size in addition a non-observable characteristics responsible for variations in growth rates for individuals of the same size is investigated. It is shown that the model can be formulated as an abstract Cauchy problem in an appropriate Banach space. We prove wellposedness of the abstract linear problem and also for a nonlinear perturbation of the model.

1991 *Mathematics Subject Classification.* 92D25, 47D03, 47H20

*Key words and phrases.* Structured populations, Sinko-Streifer model,  $C_0$ -semigroups, Lumer-Phillips theorem, semilinear abstract Cauchy problems.

### 1. Introduction

In most of the vast literature on population models (see [10] for an overview), each individual (or each individual with the same age or size) is assumed to follow the same pattern of development (e.g., growth, mortality, fecundity). However, it is well known that in addition to the observable characteristics such as age or size of individuals, non-observable genetic characteristics may often play a critical role in the development of the individuals. In the classical models proposed by F. R. Sharp, A. Lotka, A. G. McKendrick, J. W. Sinko, and W. Streifer and the

---

Research supported in part(H.T.B. and C.W.) by the U.S. National Science Foundation under the grant NSF INT-9015007 and in part (F.K.) by the Fonds zur Förderung der Wissenschaftlichen Forschung (Austria) under Grant P8146-PHY and in part by the Commission of the European Communities under the SCIENCE Program "Evolutionary Systems, Deterministic and Stochastic Evolution Equations, Control Theory and Mathematical Biology".

numerous nonlinear variations of these models, the growth rate of each individual is assumed to be a non-increasing function of the individual's age or size. As a consequence, there can be no dispersion of the density of the population in age or size (see [1] for more detailed discussions). Therefore the Sharp-Lotka-McKendrick class of models is in conflict with most of the field data collected by experimental biologists. Alternative models with diffusion mechanisms derived from random walk models (see [7] for examples) are not based on convincing biological explanations (an individual can not choose to stay young simply because there are too many older individuals). Thus, in such models some of the parameters do not lend themselves to clear biological interpretations. In [1], we first proposed that the total population should be most properly considered to be a mixture of subpopulations of individuals with specific biological characteristics assigned to a given subpopulation. For each subpopulation of individuals with identical biological characteristic, the growth pattern of the subpopulation could thus be modeled by a Sharp-Lotka-McKendrick-Sinko-Streifer type model with specific growth parameters. It was demonstrated in [1] through simulations that the solutions of the resulting heterogeneous population model exhibit dispersion similar to that observed in field data. In subsequent investigations [2], this modeling approach was given a sound theoretical foundation in the context of aggregate population probability distributions for rate parameters and parameter identification techniques for the rate distribution on the heterogeneous population model were developed. These ideas were extended in [5] where it was shown that a semigroup of linear operators on a space of vector-valued measures can be used to formulate a heterogeneous population model. A somewhat different extension of these modeling ideas was proposed in [6], where the model equation is given by

$$(1.1) \quad \begin{aligned} \partial_t u(t, x, q) &= -\partial_x(g(x; q)u(t, x; q)) - \mu(x; q)u(t, x; q), \\ t > 0, \quad x &\in (0, x_{\max}(q)), \quad q \in Q, \end{aligned}$$

$$(1.2) \quad g(0; q)u(t, 0; q) = \int_Q \int_0^{x_{\max}(\hat{q})} K(q, \hat{q}, x)u(t, x; \hat{q}) dx dm(\hat{q}), \quad t > 0, \quad q \in Q,$$

where  $q \in Q$  represents the biological characteristic of an individual,  $x_{\max}(q)$  is the maximum size attainable by individuals with characteristic  $q$ , and  $m$  is a probability measure on the space of characteristics. The model (1.1) - (1.2) is further generalized in [6] by including a nonlinear mortality rate  $\mu$  which depends on the total population

$$(1.3) \quad P(t) = \int_Q \int_0^{x_{\max}(q)} u(t, x; q) dx dm(q).$$

The global existence of solutions of (1.1) - (1.3) is established in [6] using the method of characteristics and a fixed point theorem. No results related to semi-

groups are given. In this paper, we develop a semigroup formulation for a heterogeneous population model similar to that of (1.1) – (1.3) and its nonlinear perturbation wherein the mortality depends on the total population. For the linear model, our semigroup approach is similar to that for the Sinko-Streifer type size-structured model studied in [3]. Development of such a functional analytical framework for these problems is desirable since this allows one to use powerful techniques for qualitative investigations as well as approximation studies leading to numerical solutions.

In Section 2 below, we investigate the proposed solution space. In Section 3, a semigroup formulation of the linear heterogeneous population model is given while in Section 4 we discuss a nonlinear perturbation of the linear model.

## 2. The Solution Space

As we indicated in Section 1, the set  $Q$  represents the set of possible biological characteristics vectors. It is assumed in this paper that  $Q$  is a compact metric space with a probability measure  $m(\cdot)$  defined on the Borel algebra of subsets of  $Q$ . Let  $\mathcal{H} = C(Q; L^1(0, 1))$  with a norm defined by

$$(2.1) \quad \|u\|_{\mathcal{H}} := \sup_{q \in Q} \|u(\cdot; q)\|_{L^1(0,1)} = \max_{q \in Q} \|u(\cdot; q)\|_{L^1(0,1)}.$$

It is not difficult to see that  $\mathcal{H}$  is a Banach space. We define a subspace  $H$  of  $\mathcal{H}$  as follows:

$$H = \{u \in \mathcal{H} \mid u(x; q) = 0 \text{ a.e. on } (x_{\max}(q), 1) \text{ for all } q \in Q\}$$

where  $x_{\max}(\cdot)$  is a given continuous function from  $Q$  into  $(0, 1]$ . For a given  $q \in Q$ ,  $x_{\max}(q)$  represents the maximum potential size of an individual with biological characteristic  $q$ . It is not difficult to verify that  $H$  is a closed subspace of  $\mathcal{H}$ , therefore  $H$  is also a Banach space with  $\|\cdot\|_{\mathcal{H}}$  in (2.1) as the norm. Since  $Q$  is compact, for each  $u \in H$ , the collection of  $L_1$  functions given by  $\{u(\cdot; q) \mid q \in Q\}$  is a compact subset of  $L_1(0, 1)$ . The following lemma states a few important properties of compact subsets of  $L_1(0, 1)$ .

**Lemma 2.1.** a) For any  $u \in H$  and any  $\epsilon > 0$ , there exists a constant  $\delta_0 > 0$  such that

$$\int_0^\infty |u(x + \delta; q) - u(x; q)| dx \leq \epsilon$$

for all  $\delta \in [0, \delta_0]$  and all  $q \in Q$  where  $u(\cdot; q)$  is extended by zero outside of the interval  $[0, 1]$ .

b) For any  $u \in H$  and  $\epsilon > 0$ , there exists a  $\delta_0 > 0$  such that

$$\int_A |u(x; q)| dx < \epsilon \quad \text{for all } q \in Q$$

provided  $A$  is a measurable subset of  $[0, 1]$  with  $\text{meas } A < \delta_0$ .

*Proof.* Statement (a) is a part of Kolmogoroff's compactness criterion for subsets of  $L^1(0, 1)$  (see, for instance [4], Theorem IV.8.20). To establish (b), note that from the continuity of  $u$  in  $q$  and the compactness of the space  $Q$ , we can find a family of values  $\{q_n\}_{n=1}^\infty \subset Q$  such that for all  $q \in Q$ , there exists  $q_n$  with

$$\int_0^1 |u(x; q) - u(x; q_n)| dx < \frac{\epsilon}{2}.$$

For each of the vectors  $q_n$ , there exists a constant  $\delta_n$  such that for any set  $A$  with measure  $A \leq \delta_n$ ,

$$\int_A |u(x; q_n)| dx \leq \frac{\epsilon}{2}.$$

Let  $\delta_0 = \min_{1 \leq n \leq n} \{\delta_n\}$ . Then we have that for all  $q \in Q$ ,

$$\int_A |u(x; q)| dx \leq \int_0^1 |u(x; q) - u(x; q_n)| dx + \int_A |u(x; q_n)| dx \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

□

In order to establish dissipativity of operators on  $H$ , it is useful to derive a more explicit representation of the left-sided directional derivative  $\tau_-$  of the norm  $\|\cdot\|_{\mathcal{H}}$ . Recall that for any  $u, v \in \mathcal{H}$ ,

$$\tau_-(u, v) = \lim_{h \downarrow 0} \frac{1}{h} (\|u + hv\|_{\mathcal{H}} - \|u\|_{\mathcal{H}}).$$

**Lemma 2.2.** Let  $\tau_-^{L^1}$  denote the left-hand directional derivative of the norm in  $L^1(0, 1)$  and for any  $u \in \mathcal{H}$ , let  $\Omega(u) \subset Q$  be defined as

$$\Omega(u) = \{q \in Q \mid \|u(\cdot; q)\|_{L^1(0,1)} = \|u\|_{\mathcal{H}}\}.$$

Then we have

$$\tau_-(u, v) = \inf_{q \in \Omega(u)} \tau_-^{L^1}(u(\cdot; q), v(\cdot; q)), \quad u, v \in \mathcal{H}.$$

*Proof.* We first show

$$(2.2) \quad \tau_-(u, v) \leq \inf_{q \in \Omega(u)} \tau_-^{L^1}(u(\cdot; q), v(\cdot; q)).$$

Consider a sequence  $\epsilon_n$  of negative real numbers such that  $\lim_{n \rightarrow \infty} \epsilon_n = 0$ . Let  $q_0$  and  $\{q_n\}$  be arbitrarily chosen such that  $q_0 \in \Omega(u)$ ,  $q_n \in \Omega(u + \epsilon_n v)$ . As a result,

$$\begin{aligned} \frac{1}{\epsilon_n} (\|u + \epsilon_n v\|_{\mathcal{H}} - \|u\|_{\mathcal{H}}) &= \frac{1}{\epsilon_n} (\|u(\cdot; q_n) + \epsilon_n v(\cdot; q_n)\|_{L^1(0,1)} - \|u(\cdot; q_0)\|_{L^1(0,1)}) \\ &\leq \frac{1}{\epsilon_n} (\|u(\cdot; q_0) + \epsilon_n v(\cdot; q_0)\|_{L^1(0,1)} - \|u(\cdot; q_0)\|_{L^1(0,1)}). \end{aligned}$$

Taking the limit as  $n \rightarrow \infty$ , we obtain (2.2). On the other hand, for any sequence of negative real numbers  $\epsilon_n$  with  $\lim_{n \rightarrow \infty} \epsilon_n = 0$ , let  $q_n \in \Omega(u + \epsilon_n v)$ . By compactness

of  $Q$ , there exists a subsequence, again denoted by  $q_n$  such that  $\lim_{n \rightarrow \infty} q_n = q_0$  for some  $q_0 \in Q$ . For any  $q \in Q$ , we have

$$\|u(\cdot; q) + \epsilon_n v(\cdot; q)\|_{L^1(0,1)} \leq \|u(\cdot; q_n) + \epsilon_n v(\cdot; q_n)\|_{L^1(0,1)} \leq \|u(\cdot; q_n)\|_{L^1(0,1)} + |\epsilon_n| \|v\|_{\mathcal{H}}.$$

By taking the limit as  $n \rightarrow \infty$ , we obtain

$$\|u(\cdot; q)\|_{L^1(0,1)} \leq \|u(\cdot; q_0)\|_{L^1(0,1)}$$

and hence  $q_0 \in \Omega(u)$ . But

$$\begin{aligned} \frac{1}{\epsilon_n} (\|u + \epsilon_n v\|_{\mathcal{H}} - \|u\|_{\mathcal{H}}) &= \frac{1}{\epsilon_n} \left( \|u(\cdot; q_n) + \epsilon_n v(\cdot; q_n)\|_{L^1(0,1)} - \|u(\cdot; q_0)\|_{L^1(0,1)} \right) \\ &\geq \frac{1}{\epsilon_n} \left( \|u(\cdot; q_n) + \epsilon_n v(\cdot; q_n)\|_{L^1(0,1)} - \|u(\cdot; q_n)\|_{L^1(0,1)} \right). \end{aligned}$$

The right side of the above inequality is equal to

$$\frac{1}{\epsilon_n} \int_0^1 (|u(x; q_n) + \epsilon_n v(x; q_n)| - |u(x; q_n)|) dx.$$

Since  $\{u(\cdot; q_n)\}$  and  $\{v(\cdot; q_n)\}$  converge to  $u(\cdot; q_0)$ ,  $v(\cdot; q_0)$  in  $L^1(0,1)$ , respectively, as  $n \rightarrow \infty$ , we can find a subsequence, again denoted by  $q_n$  such that  $\{u(\cdot; q_n)\}$ ,  $\{v(\cdot; q_n)\}$  converge to  $u(\cdot; q_0)$ ,  $v(\cdot; q_0)$  almost everywhere. Let  $\Sigma_1(u(\cdot; q_0))$  and  $\Sigma_0(u(\cdot; q_0))$  be two subsets of the interval  $[0,1]$ , defined by

$$\begin{aligned} \Sigma_1(u(\cdot; q_0)) &= \{x \in [0,1] \mid u(x; q_0) \neq 0\}, \\ \Sigma_0(u(\cdot; q_0)) &= \{x \in [0,1] \mid u(x; q_0) = 0\}. \end{aligned}$$

Then we can argue that

$$\lim_{n \rightarrow \infty} \frac{1}{\epsilon_n} (|u(x; q_n) + \epsilon_n v(x; q_n)| - |u(x; q_n)|) = \text{sign}(u(x; q_0)) v(x; q_0)$$

almost everywhere on  $\Sigma_1(u(\cdot; q_0))$ . Hence by dominated convergence, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{\Sigma_1(u(\cdot; q_0))} \frac{1}{\epsilon_n} (|u(x; q_n) + \epsilon_n v(x; q_n)| - |u(x; q_n)|) dx \\ = \int_{\Sigma_1(u(\cdot; q_0))} \text{sign}(u(x; q_0)) v(x; q_0) dx. \end{aligned}$$

On the subset  $\Sigma_0(u(\cdot; q_0))$ , we have

$$\frac{1}{\epsilon_n} (|u(x; q_n) + \epsilon_n v(x; q_n)| - |u(x; q_n)|) \geq -|v(x; q_n)|.$$

As a result, we find

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{\epsilon_n} \int_{\Sigma_0(u(\cdot; q_0))} (|u(x; q_n) + \epsilon_n v(x; q_n)| - |u(x; q_n)|) dx \\ \geq - \int_{\Sigma_0(u(x; q_0))} |v(x; q_0)| dx. \end{aligned}$$

Combining the above results, we obtain

$$\begin{aligned} \tau^-(u, v) &= \lim_{n \rightarrow \infty} \frac{1}{\epsilon_n} (\|u + \epsilon_n v\|_H - \|u\|_H) \\ &\geq \int_{\Sigma_0(u(\cdot; q_0))} \text{sign}(u(x; q_0)) v(x; q_0) dx - \int_{\Sigma_0(u(x; q_0))} |v(x; q_0)| dx \\ &= \tau^{L^1}(u(\cdot; q_0), v(\cdot; q_0)). \end{aligned}$$

□

### 3. Well-Posedness of the Linear Model

In this section, we rewrite equations (1.1)–(1.2) as an evolution equation in  $H$  of the form

$$\frac{d}{dt} u(t) = Au(t), \quad t \geq 0.$$

To this end, we make some standing assumptions on the growth, mortality, and fecundity functions in (1.1)–(1.2). We assume:

- (A1) The growth rate function satisfies  $g \in C^1(Q, W^{1,\infty}(0, 1))$  and for each  $q \in Q$  we have  $g(x; q) > 0$  on  $[0, x_{\max}(q))$  and  $g(x; q) = 0$  for  $x \in [x_{\max}(q), 1]$ .
- (A2) For each  $q \in Q$ ,  $\mu(\cdot; q) \in L_{\text{loc}}^\infty(0, x_{\max}(q))$  with  $\mu(\cdot; q) \geq 0$  a.e. on  $[0, 1]$  and  $\mu(\cdot; q) > 0$  on  $[x_{\max}(q), 1]$ . For any  $\epsilon \in (0, x_{\max}(q))$ , we define  $\mu_\epsilon$  by

$$\mu_\epsilon(x; q) = \begin{cases} \mu(x; q) & \text{for } x \in (0, x_{\max}(q) - \epsilon), \\ 0 & \text{otherwise,} \end{cases}$$

and assume that  $\mu_\epsilon \in C(Q, L^\infty(0, 1))$  for  $0 < \epsilon < \min_{q \in Q} x_{\max}(q)$ .

- (A3) The function  $K$  is in  $C(Q \times Q; L^\infty(0, 1))$  with  $K(q, \tilde{q}, \cdot) = 0$  a.e. on  $[x_{\max}(q), 1]$ .

We shall define an unbounded linear operator  $A$  and show that  $A$  is the infinitesimal generator of a  $C_0$ -semigroup of bounded linear operators on  $H$ . The domain  $\mathcal{D}(A)$  of  $A$  is the set of all functions  $u \in H$  satisfying the following conditions:

- (i) For all  $q \in Q$ ,  $u(\cdot; q)$  is locally absolutely continuous on  $[0, x_{\max}(q))$ ;
- (ii) The function defined by  $q \rightarrow \partial_x(g(x; q)u(x; q)) + \mu(x; q)u(x; q)$  for  $x \in [0, x_{\max}(q))$ ,  $q \in Q$ , is an element in  $H$ ;

(iii) For any  $q \in Q$ ,

$$g(0; q)u(0; q) = \int_Q \int_0^{x_{\max}(\hat{q})} K(q, \hat{q}, x)u(x; \hat{q}) dx d\mu(\hat{q});$$

(iv) For any  $q \in Q$ ,  $\lim_{x \uparrow x_{\max}(q)} g(x; q)u(x; q) = 0$ .

For  $u \in \mathcal{D}(A)$ , we define

$$(Au)(x; q) = -\partial_x(g(x; q)u(x; q)) - \mu(x; q)u(x; q), \quad 0 < x < x_{\max}(q), \quad q \in Q.$$

The main result of this section is stated as follows.

**Theorem 3.1.** *Under assumptions (A1)–(A3), the operator  $A$  defined above is the infinitesimal generator of a  $C_0$ -semigroup  $\{S(t), t \geq 0\}$  of bounded linear operators on  $H$ .*

The proof of this theorem is based on the use of the Lumer-Phillips theorem which requires that the infinitesimal generator  $A$  satisfies the following:

- (1) The domain of  $A$  must be dense in  $H$ ;
- (2) There exists a real number  $\omega$  such that  $A - \omega I$  is  $m$ -dissipative, i.e., the range of  $A - \omega I$  is  $H$  and  $A - \omega I$  is dissipative.

**3.1. Density of the domain  $\mathcal{D}(A)$ .** Consider the following subset of  $H$ :

$$H_\epsilon = \{u \in H \mid u(x; q) = 0 \text{ for } x \in (x_{\max}(q) - \epsilon, 1) \text{ for all } q \in Q\}.$$

Let  $H_0 = \bigcup_{\epsilon > 0} H_\epsilon$ . From Lemma 2.1, we conclude that  $H_0$  is dense in  $H$ . We first want to show that a subset of  $H_0$  that contains functions differentiable in  $x$  is also dense in  $H$ . Consider a sequence  $\{L_k\}$  of bounded linear operators  $L_k : L^1(0, 1) \rightarrow C^1(0, 1)$  defined by

$$(L_k f)(x) = \int_0^1 f(y) \lambda_k(x - y) dy$$

where

$$\lambda_k(x) = \begin{cases} \theta_k e^{-(1 - (kx)^2)^{-1}} & \text{for } x \in (-1/k, 1/k), \\ 0 & \text{elsewhere.} \end{cases}$$

The constant  $\theta_k$  is chosen such that

$$\int_{-\infty}^{+\infty} \lambda_k(x) dx = \int_{-1/k}^{1/k} \lambda_k(x) dx = 1.$$

The following properties of  $L_k$  are useful.

**Lemma 3.1.** (i) *For any  $f \in L^1(0, 1)$  such that  $f(x) = 0$  for all  $x > \delta$  with a  $\delta < 1$ , we have  $(L_k f)(x) = 0$  for all  $x > \delta + 1/k$  whenever  $\delta + 1/k < 1$ .*

(ii)  $\lim_{k \rightarrow \infty} \|f - L_k f\|_{L^1(0, 1)} = 0$ .

*Proof.* (i) is obvious. It is sufficient to show that (ii) holds for all continuous functions. In fact,

$$\begin{aligned}\|f - L_k f\|_{L^1(0,1)} &= \int_0^1 \left| \int_0^1 (f(x) - f(y)) \lambda_k(x-y) dy \right| dx \\ &= \int_0^1 \left| \int_{x-1/k}^{x+1/k} (f(x) - f(y)) \lambda_k(x-y) dy \right| dx.\end{aligned}$$

It is easy to see that

$$\left| \int_{x-1/k}^{x+1/k} (f(x) - f(y)) \lambda_k(x-y) dy \right| \leq \sup_{y \in (x-1/k, x+1/k)} |f(x) - f(y)| \leq 2 \sup_{x \in (0,1)} |f(x)|,$$

and therefore, since  $f$  is continuous,

$$\lim_{k \rightarrow \infty} \left| \int_{x-1/k}^{x+1/k} (f(x) - f(y)) \lambda_k(x-y) dy \right| = 0 \quad \text{for all } x \in (0,1).$$

By the dominated convergence theorem, we obtain the desired result (ii).  $\square$

Let  $V_\epsilon$  be the subset of  $H_\epsilon$  defined by

$$V_\epsilon = \{u \in H_\epsilon \mid u \in C(Q; C^1(0,1))\}$$

and let  $V_0 = \cup_{\epsilon > 0} V_\epsilon$ . Then we have

**Lemma 3.2.**  $V_0$  is a dense subset of  $H$ .

*Proof.* It is sufficient to show that every element  $u \in H_0$  can be approximated by functions in  $V_0$ . Consider an element  $u \in H_0$  and define  $u_k$  as follows,

$$u_k(\cdot; q) = L_k u(\cdot; q), \quad \text{for all } q \in Q.$$

We want to show that  $u_k \in V_0$  for all  $k$  sufficiently large and  $\lim_{k \rightarrow \infty} \|u_k - u\|_H = 0$ . Since  $u \in H_0$ , there exists  $\epsilon > 0$  such that  $u \in H_\epsilon$ . For all  $k > 2/\epsilon$ ,  $u_k \in H_{\frac{\epsilon}{2}}$ . On the other hand, for all  $q \in Q$ ,  $u_k(\cdot; q) \in C^1(0,1)$ . Moreover, let  $C'_k$  be a constant such that

$$\|L_k\|_{\mathcal{L}(L^1(0,1), C^1(0,1))} \leq C'_k.$$

Then,

$$\|u_k(\cdot; q) - u_k(\cdot; \hat{q})\|_{C^1(0,1)} \leq C'_k \|u(\cdot; q) - u(\cdot; \hat{q})\|_{L^1(0,1)}.$$

Since  $u \in C(Q, L^1(0,1))$ , we conclude that

$$\lim_{d(q, \hat{q}) \rightarrow 0} \|u_k(\cdot; q) - u_k(\cdot; \hat{q})\|_{C^1(0,1)} = 0,$$

where  $d$  is the metric for  $Q$ .

As a result, for  $k$  sufficiently large, we have  $u_k \in V_{\frac{\epsilon}{2}}$ . Using Lemma 3.1, we find

$$\lim_{k \rightarrow \infty} \|u_k(\cdot; q) - u(\cdot; q)\|_{L^1(0,1)} = 0, \quad \text{for all } q \in Q.$$



On the other hand, by compactness of  $Q$  and the uniform boundedness of operators  $L_k : L^1(0, 1) \rightarrow L^1(0, 1)$ , we conclude that

$$\lim_{k \rightarrow \infty} \sup_{q \in Q} \|u_k(\cdot; q) - u(\cdot; q)\|_{L^1(0,1)} = 0,$$

or  $u_k \rightarrow u$  in  $H$ , which establishes the density of  $V_0$  in  $H$ .  $\square$

Finally, for a given element  $u \in V_0$ , we shall construct a sequence of functions  $u_\epsilon$  in  $\mathcal{D}(A)$  such that  $u_\epsilon$  converges to  $u$  in  $H$ . The basic idea is to modify the function  $u(\cdot; q)$  in the neighborhood of  $x = 0$  so that the resulting function satisfies the boundary condition at  $x = 0$ . Let  $\epsilon > 0$  be a small real value. We define the modified function  $u_\epsilon$  by

$$u_\epsilon(x; q) = \begin{cases} (1 - x/\epsilon)v_\epsilon(q) + (x/\epsilon)u(\epsilon; q) & \text{for } 0 \leq x \leq \epsilon, \\ u(x; q) & \text{for } \epsilon < x \leq 1. \end{cases}$$

where the function  $v_\epsilon$  should be chosen such that  $u_\epsilon \in \mathcal{D}(A)$ . In particular,  $u_\epsilon$  should satisfy

$$g(0; q)u_\epsilon(0; q) = \int_Q \int_0^{x_{\max}(q)} K(q, \hat{q}, x) u_\epsilon(x) dx dm(\hat{q}).$$

Using the definition of  $u_\epsilon$ , we must be able to choose the function  $v = v_\epsilon$  to satisfy  $\mathcal{T}_\epsilon v = v$  where the mapping  $\mathcal{T}_\epsilon : C(Q) \rightarrow C(Q)$  is defined as

$$\begin{aligned} (\mathcal{T}_\epsilon v)(q) &= \int_Q \int_0^\epsilon \left(1 - \frac{x}{\epsilon}\right) \frac{K(q, \hat{q}, x)}{g(0; q)} v(\hat{q}) dx dm(\hat{q}) \\ &\quad + \int_Q \int_0^\epsilon \frac{K(q, \hat{q}, x)}{g(0; q)} \frac{x}{\epsilon} u(\epsilon; \hat{q}) dx dm(\hat{q}) \\ &\quad + \int_Q \int_\epsilon^{x_{\max}(q)} \frac{K(q, \hat{q}, x)}{g(0; q)} u(x; \hat{q}) dx dm(\hat{q}). \end{aligned}$$

To see that this is possible we argue:

**Lemma 3.3.** *For  $\epsilon$  sufficiently small, the mapping  $\mathcal{T}_\epsilon$  is a contraction from  $C(Q)$  into  $C(Q)$ .*

*Proof.* The continuity of  $\mathcal{T}_\epsilon v$  for  $v \in C(Q)$  and  $u \in V_0$  follows immediately from assumptions (A3) and (A1). On the other hand, for any  $v, \hat{v} \in C(Q)$ , we have

$$(\mathcal{T}_\epsilon v - \mathcal{T}_\epsilon \hat{v})(q) = \int_Q \int_0^\epsilon \left(1 - \frac{x}{\epsilon}\right) \frac{K(q, \hat{q}, x)}{g(0; q)} (v(\hat{q}) - \hat{v}(\hat{q})) dx dm(\hat{q}).$$

As a result,

$$|(\mathcal{T}_\epsilon v - \mathcal{T}_\epsilon \hat{v})(q)| \leq \sup_{\hat{q} \in Q} |v(\hat{q}) - \hat{v}(\hat{q})| \int_Q \int_0^\epsilon \left(1 - \frac{x}{\epsilon}\right) \frac{K(q, \hat{q}, x)}{g(0; q)} dx dm(\hat{q}).$$

where

$$\lim_{\epsilon \rightarrow 0} \int_Q \int_0^\epsilon \left(1 - \frac{x}{\epsilon}\right) \frac{K(q, \hat{q}, x)}{g(0, q)} dx dm(\hat{q}) = 0.$$

Therefore, for all  $\epsilon$  sufficiently small, there exists  $\gamma_\epsilon < 1$  such that

$$\sup_{q \in Q} |(\mathcal{T}_\epsilon v - \mathcal{T}_\epsilon \hat{v})(q)| \leq \gamma_\epsilon \sup_{q \in Q} |(v - \hat{v})(q)|,$$

which proves the lemma.  $\square$

Using the contraction mapping theorem, there exists a unique fixed point  $v_\epsilon$  for the mapping  $\mathcal{T}_\epsilon$  in  $C(Q)$ . Since  $u \in V_0$ , it is not difficult to see that  $u_\epsilon$  is piecewise continuously differentiable, and there exists  $\delta > 0$  such that  $u_\epsilon(x; q) = 0$  for  $x \in [x_{\max}(q) - \delta, 1]$ . Thus  $u_\epsilon$  satisfies all the conditions required of an element of  $\mathcal{D}(A)$ . Moreover, we can also show that  $u_\epsilon$  converges to  $u$  in the  $H$ -topology as  $\epsilon \rightarrow 0$ .

**Lemma 3.4.** *Let  $u_\epsilon$  be defined as above. Then*

$$\lim_{\epsilon \rightarrow 0} \|u_\epsilon - u\|_H = 0.$$

*Proof.* Considering the term  $u_\epsilon(\cdot; q) - u(\cdot; q)$ , we have the estimate

$$\begin{aligned} \|u_\epsilon(\cdot; q) - u(\cdot; q)\|_{L^1(0,1)} &\leq \int_0^\epsilon |u(x; q)| dx + \frac{1}{\epsilon} \int_0^\epsilon x |u(\epsilon; q)| dx + \int_0^\epsilon \left(1 - \frac{x}{\epsilon}\right) |v_\epsilon(q)| dx \\ &\leq \sup_{q \in Q} \int_0^\epsilon |u(x; \hat{q})| dx + \frac{\epsilon}{2} \sup_{q \in Q} \sup_{0 \leq x \leq 1} |u(x; \hat{q})| + \epsilon \sup_{q \in Q} |v_\epsilon(q)|. \end{aligned}$$

From the definition of  $v_\epsilon$ , we have

$$\sup_{q \in Q} |v_\epsilon(q)| \leq \frac{C_0 + \epsilon C_2}{1 - \epsilon C_1}$$

where

$$\begin{aligned} C_1 &= \sup_{q \in Q} \int_Q \frac{\|K(q, \hat{q}, \cdot)\|_{L^\infty(0,1)}}{g(0; q)} dm(\hat{q}) \\ C_2 &= \sup_{q \in Q} \int_Q \frac{\|K(q, \hat{q}, \cdot) u(\cdot; \hat{q})\|_{L^\infty(0,1)}}{g(0; q)} dm(\hat{q}) \\ C_0 &= \sup_{q \in Q} \int_Q \int_0^{x_{\max}(\hat{q})} \left| \frac{K(q, \hat{q}, x)}{g(0; q)} u(x; \hat{q}) \right| dx dm(\hat{q}). \end{aligned}$$

Therefore

$$\sup_{q \in Q} |v_\epsilon(q)| \leq C \quad \text{for all } \epsilon > 0 \text{ sufficiently small.}$$

As a result, we conclude that

$$\lim_{\epsilon \rightarrow 0} \sup_{q \in Q} \|u_\epsilon(\cdot; q) - u(\cdot; q)\|_{L^1(0,1)} = 0.$$

□

Combining the above sequence of lemmas and results, we can conclude that  $\mathcal{D}(A)$  is dense in  $H$ .

**3.2. Dissipativity of the operator  $A$ .** Let us recall that the operator  $A - \lambda I$ ,  $\lambda \in \mathbb{R}$ , is dissipative if and only if

$$\tau_-(u, Au) \leq \lambda \|u\|_{\mathcal{H}} \quad \text{for all } u \in \mathcal{D}(A).$$

Using Lemma 2.2, we have

$$\tau_-(u, Au) = \inf_{q \in \Omega(u)} \tau_-^{L^1}(u(\cdot; q), (Au)(\cdot; q)), \quad u \in \mathcal{D}(A),$$

where  $\Omega(u) = \{q \in Q \mid \|u(\cdot; q)\|_{L^1(0,1)} = \|u\|_{\mathcal{H}}\}$ . By the same computations as presented in Lemma 2.1 of [3], we obtain immediately

$$\begin{aligned} \tau_-^{L^1}(u(\cdot; q), (Au)(\cdot; q)) &\leq g(0; q) |u(\cdot; q)| = \left| \int_Q \int_0^{r_{\max}(\hat{q})} K(q, \hat{q}, x) u(x; \hat{q}) dx dm(\hat{q}) \right| \\ &\leq k_0 \|u\|_{\mathcal{H}} \end{aligned}$$

for  $q \in Q$ ,  $u \in \mathcal{D}(A)$ , where  $k_0 = \sup_{q, \hat{q} \in Q} \|K(q, \hat{q}, \cdot)\|_{L^\infty(0,1)}$ . Thus  $A - k_0 I$  is dissipative on  $H$ .

**3.3. The range of the operator  $A - \lambda I$ .** Since the range of the operator  $A - \lambda I$  and the range of a nonlinear perturbation of  $A$  can be found using similar arguments, we will try to establish a result that is useful in both cases. We first define a convex functional  $p$  on  $H$  by

$$p(u) = \|u\|_{\mathcal{H}}.$$

Furthermore, we define the nonlinear mapping  $F: H \rightarrow H$  by

$$(Fu)(x; q) = -\mu_0(x; q) P(u) u(x; q),$$

where  $\mu_0 \in C(Q; L^\infty(0, 1))$ ,  $\mu_0(x; q) \geq 0$  for all  $x \in (0, 1)$  and  $q \in Q$  and

$$P(u) = \int_Q \int_0^{r_{\max}(q)} |u(x; q)| dx dm(q).$$

We would like to show the following result.

**Theorem 3.2.** *Let the subset  $C(r)$  of  $H$  be defined by  $C(r) = \{u \in H \mid p(u) \leq r\}$ . Then under assumptions (A1)–(A3) and those above on  $\mu_0$ , one can find a constant  $a > 0$  such that the following holds:*

*Given  $r \geq 0$ , there exists a number  $\lambda_0 \equiv \lambda_0(r) \in (0, 1/a)$  such that to each  $u \in C(r)$  and each  $\lambda \in (0, \lambda_0)$ , there corresponds  $u_\lambda \in \mathcal{D}(A)$  satisfying*

- (a)  $u_\lambda - \lambda(A + F)u_\lambda = u$ ,  
 (b)  $p(u_\lambda) \leq (1 - a\lambda)^{-1}p(u)$ .

We observe that when  $\mu_0 \equiv 0$ , statement (a) implies that the range of the operator  $A - \lambda I$  is  $H$  for all  $\lambda > a$ . On the other hand, (a) is equivalent to

$$\frac{1}{\lambda} u_\lambda(x; q) + \partial_x(g(x; q)u_\lambda(x; q)) + (\mu(x; q) + \mu_0(x; q)P(u_\lambda))u_\lambda(x; q) = \frac{1}{\lambda} u(x; q)$$

for  $x \in (0, x_{\max}(q))$ ,  $q \in Q$ . Combining this with the boundary condition (iii) in the definition of  $\mathcal{D}(A)$ , we find that  $u_\lambda$  must satisfy

$$(3.1) \quad \begin{aligned} u_\lambda(x; q) = & \frac{E(x, 0; q, u_\lambda)}{g(x; q)} \int_Q \int_0^{x_{\max}(\hat{q})} K(q, \hat{q}, y) u_\lambda(y, \hat{q}) dy dm(\hat{q}) \\ & + \frac{1}{\lambda g(x; q)} \int_0^x E(x, z; q, u_\lambda) u(z; q) dz \end{aligned}$$

where

$$E(x, z; q, u_\lambda) = \exp\left(-\int_z^x \frac{1/\lambda + \mu(\zeta; q) + \mu_0(\zeta; q)P(u_\lambda)}{g(\zeta; q)} d\zeta\right).$$

The proof of the above theorem consists of establishing the existence of the solution of (3.1) with property (b) for a certain nonnegative constant  $a$ . Indeed, we will establish (b) with the constant  $a$  given in the following lemma.

**Lemma 3.5.** *If  $u_\lambda \in H$  is a solution of the integral equation (3.1), then*

$$\|u_\lambda\|_{\mathcal{H}} \leq (1 - a\lambda)^{-1} \|u\|_{\mathcal{H}},$$

where

$$a = \sup_{\|u\|_{\mathcal{H}} \leq 1, q \in Q} \left| \int_Q \int_0^{x_{\max}} K(q, \hat{q}, x) u(x; \hat{q}) dx dm(\hat{q}) \right|.$$

*Proof.* From (3.1), we have

$$|u_\lambda(x; q)| \leq \frac{E(x, 0; q, u_\lambda)}{g(x; q)} a \|u_\lambda\|_{\mathcal{H}} + \left| \frac{1}{\lambda g(x; q)} \int_0^x E(x, z; q, u_\lambda) u(z; q) dz \right|.$$

If we integrate both sides of the above inequality, we obtain

$$\begin{aligned} \int_0^1 |u_\lambda(x; q)| dx & \leq \int_0^1 \frac{E(x, 0; q, u_\lambda)}{g(x; q)} dx a \|u_\lambda\|_{\mathcal{H}} \\ & \quad + \int_0^1 \frac{1}{\lambda g(x; q)} \int_0^x E(x, z; q, u_\lambda) |u(z; q)| dz dx. \end{aligned}$$

But

$$\int_0^1 \frac{E(x, 0; q, u_\lambda)}{g(x; q)} dx \leq \int_0^1 \frac{1}{g(x; q)} \exp\left(-\int_0^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) dx \leq \lambda.$$

and

$$\begin{aligned} \int_0^1 \frac{1}{\lambda g(x; q)} \int_0^x E(x, z; q, u_\lambda) |u(z; q)| dz dx \\ = \int_0^1 |u(z; q)| \int_z^1 \frac{1}{\lambda g(x; q)} E(x, z; q, \mu_\lambda) dx dz. \end{aligned}$$

Arguing as above, we find

$$\int_z^1 \frac{1}{\lambda g(x; q)} E(x, z; q, \mu_\lambda) dx \leq \int_z^1 \frac{1}{\lambda g(x; q)} \exp\left(-\int_z^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) dx \leq 1.$$

Therefore,

$$\int_0^1 \frac{1}{\lambda g(x; q)} \int_0^x E(x, z; q, u_\lambda) |u(z; q)| dz dx \leq \int_0^1 |u(x; q)| dx.$$

As a result, we have

$$\|u_\lambda\|_{\mathcal{H}} \leq \lambda a \|u_\lambda\|_{\mathcal{H}} + \|u\|_{\mathcal{H}},$$

or

$$\|u_\lambda\|_{\mathcal{H}} \leq \frac{\|u\|_{\mathcal{H}}}{1 - \lambda a},$$

for  $\lambda < 1/a$ .  $\square$

The above lemma provides an *a priori* estimate for the solution of (3.1). To prove Theorem 3.2, it is sufficient to show that (a) has a solution in  $\mathcal{D}(A)$ . We define a mapping  $T : H \rightarrow H$  by

$$\begin{aligned} (Tv)(x; q) &= \frac{1}{g(x; q)} E(x, 0; q, v) \int_Q \int_0^{x_{\max}(q)} K(q, \dot{q}, \zeta) v(\zeta; \dot{q}) d\zeta dm(\dot{q}) \\ &\quad + \frac{1}{\lambda g(x; q)} \int_0^x E(x, z; q, v) u(z; q) dz, \end{aligned}$$

where

$$\begin{aligned} E(x, z; q, v) &= \exp\left(-\int_z^x \frac{1/\lambda + \mu(\zeta; q) + \mu_0(\zeta; q)P(v)}{g(\zeta; q)} d\zeta\right), \\ P(v) &= \int_Q \int_0^{x_{\max}(q)} |v(x; q)| dx dm(q). \end{aligned}$$

**Lemma 3.6.** *For any given  $r > 0$  and  $u \in C(r)$ , there exists  $\lambda_0 \in (0, 1/a)$  such that, for all  $\lambda \leq \lambda_0$ , the mapping  $T$  is a contraction on  $C(r/(1 - \lambda_0 a))$ .*

*Proof.* For any  $v \in H$  from the definition of  $Tv$ , it is not difficult to see that  $(Tv)(\cdot; q) \in L^1(0, 1)$  for all  $q \in Q$ . Moreover, using the continuity properties of the functions  $v$ ,  $K$ ,  $\mu$  and  $g$ , one can establish that  $Tv \in H$ .

Now consider two elements  $v, \hat{v} \in H$ . We assume without loss of generality that  $P(v) \leq P(\hat{v})$ . Then we have

$$\begin{aligned} & |(Tv)(x; q) - (T\hat{v})(x; q)| \\ & \leq \frac{1}{g(x; q)} E(x, 0; q, v) \int_Q \int_0^{x_{\max}(\hat{q})} K(q, \hat{q}, z) |v(z; \hat{q}) - \hat{v}(z; \hat{q})| dz dm(\hat{q}) \\ & \quad + \frac{1}{g(x; q)} (E(x, 0; q, v) - E(x, 0; q, \hat{v})) \int_Q \int_0^{x_{\max}(q)} K(q, \hat{q}, z) |\hat{v}(z; q)| dz dm(\hat{q}) \\ & \quad + \frac{1}{\lambda g(x; q)} \int_0^x |u(z; q)| (E(x, z; q, v) - E(x, z; q, \hat{v})) dz. \end{aligned}$$

By integrating both sides we obtain

$$\begin{aligned} & \int_0^{x_{\max}(q)} |(Tv)(x; q) - (T\hat{v})(x; q)| dx \\ & \leq \int_0^{x_{\max}(q)} \frac{1}{g(x; q)} E(x, 0; q, v) \int_Q \int_0^{x_{\max}(\hat{q})} K(q, \hat{q}, z) |v - \hat{v}| dz dm(\hat{q}) dx \\ & \quad + \int_0^{x_{\max}(q)} \frac{1}{g(x; q)} (E(x, 0; q, v) - E(x, 0; q, \hat{v})) \int_Q \int_0^{x_{\max}(\hat{q})} K(q, \hat{q}, z) |\hat{v}(z; q)| dz dm(\hat{q}) dx \\ & \quad + \int_0^{x_{\max}(q)} \frac{1}{\lambda g(x; q)} \int_0^x |u(z; q)| (E(x, z; q, v) - E(x, z; q, \hat{v})) dz dx \\ & =: I_1 + I_2 + I_3. \end{aligned}$$

The term  $I_1$  can be estimated by

$$\begin{aligned} & \int_0^{x_{\max}(q)} \frac{1}{g(x; q)} E(x, 0; q, v) \int_Q \int_0^{x_{\max}(\hat{q})} K(q, \hat{q}, z) |(v - \hat{v})(z; \hat{q})| dz dm(\hat{q}) dx \\ & \leq k_0 \|v - \hat{v}\|_{\mathcal{H}} \int_0^{x_{\max}(q)} \frac{1}{g(x; q)} \exp\left(-\int_0^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) dx = k_0 \lambda \|v - \hat{v}\|_{\mathcal{H}}. \end{aligned}$$

The term  $I_2$  can be estimated by

$$\begin{aligned}
& \int_0^{x_{\max}(q)} \frac{1}{g(x; q)} (E(x, 0; q, v) - E(x, 0; q, \hat{v})) \int_Q \int_0^{x_{\max}(\hat{q})} K(q, \hat{q}, z) |\hat{v}(z; q)| dz dm(\hat{q}) dx \\
& \leq k_0 \|\hat{v}\|_{\mathcal{H}} \left( \int_0^{x_{\max}(q)} \frac{1}{g(x; q)} \exp\left(-\int_0^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) \left( \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(v) d\zeta\right) \right. \right. \\
& \quad \left. \left. - \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(\hat{v}) d\zeta\right) \right) dx \right. \\
& = -k_0 \|\hat{v}\|_{\mathcal{H}} \lambda \left( \exp\left(-\int_0^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) \right. \\
& \quad \times \left( \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(v) d\zeta\right) - \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(\hat{v}) d\zeta\right) \right) \Big|_{x=0}^{x=x_{\max}(q)} \\
& \quad + \lambda \int_0^{x_{\max}(q)} \exp\left(-\int_0^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) \frac{\mu_0(x; q)}{g(x; q)} \\
& \quad \times \left( (P(v) - P(\hat{v})) \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(v) d\zeta\right) \right. \\
& \quad \left. + P(\hat{v}) \left( \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(v) d\zeta\right) - \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(\hat{v}) d\zeta\right) \right) \right) dx.
\end{aligned}$$

The first term in the above expression is equal to zero (assumption (A1) implies  $\int_0^{x_{\max}(q)} \frac{d\zeta}{g(\zeta; q)} = \infty$ ). On the other hand, we have

$$\left| \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(v) d\zeta\right) - \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(\hat{v}) d\zeta\right) \right| \leq \theta |P(v) - P(\hat{v})|$$

for some constant  $\theta$ . As a result, we have

$$\begin{aligned}
& \lambda \int_0^{x_{\max}(q)} \exp\left(-\int_0^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) \frac{\mu_0(x; q)}{g(x; q)} \left( (P(v) - P(\hat{v})) \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(v) d\zeta\right) \right. \\
& \quad \left. + P(\hat{v}) \left( \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(v) d\zeta\right) - \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(\hat{v}) d\zeta\right) \right) \right) dx \\
& \leq \lambda^2 \|\hat{v} - v\|_{\mathcal{H}} \int_0^{x_{\max}(q)} \exp\left(-\int_0^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) \frac{\mu_0(x; q)}{\lambda g(x; q)} dx \\
& \quad + \lambda^2 \theta \|\hat{v} - v\|_{\mathcal{H}} P(\hat{v}) \int_0^{x_{\max}(q)} \frac{\mu_0(x; q)}{\lambda g(x; q)} \exp\left(-\int_0^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) dx.
\end{aligned}$$

Moreover, if  $\mu_0 \in C(Q, L^\infty(0, 1))$ , we have

$$\int_0^{x_{\max}(q)} \frac{\mu_0(x; q)}{\lambda g(x; q)} \exp\left(-\int_0^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) dx \leq \|\mu_0\|_{C(Q, L^\infty(0, 1))}.$$

The term  $I_3$  can be estimated in a similar manner by

$$\begin{aligned} & \int_0^{x_{\max}(q)} \frac{1}{\lambda g(x; q)} \int_0^x |u(z; q)| (E(x, z; q, v) - E(x, z; q, \hat{v})) dz dx \\ & \leq \int_0^{x_{\max}(q)} |u(z; q)| \int_z^{x_{\max}(q)} \frac{1}{\lambda g(x; q)} \exp\left(-\int_z^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) \\ & \quad \times \left( \exp\left(-\int_0^x \frac{\mu_0}{g} P(v) d\zeta\right) - \exp\left(-\int_z^x \frac{\mu_0}{g} P(\hat{v}) d\zeta\right) \right) dx dz. \end{aligned}$$

From the estimate of  $I_2$ , we have

$$\begin{aligned} & \int_z^{x_{\max}(q)} \frac{1}{\lambda g(x; q)} \exp\left(-\int_z^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) \\ & \quad \times \left( \exp\left(-\int_0^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(v) d\zeta\right) - \exp\left(-\int_z^x \frac{\mu_0(\zeta; q)}{g(\zeta; q)} P(\hat{v}) d\zeta\right) \right) dx \\ & \leq \lambda \|v - \hat{v}\|_{\mathcal{H}} (1 + \theta) \int_z^{x_{\max}(q)} \frac{\mu_0(x; q)}{\lambda g(x; q)} \exp\left(-\int_z^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) dx. \end{aligned}$$

As a result, we have

$$\begin{aligned} I_3 & \leq \lambda \|v - \hat{v}\| (1 + \theta) \|u\|_{\mathcal{H}} \int_0^{x_{\max}(q)} \frac{\mu_0(x; q)}{\lambda g(x; q)} \exp\left(-\int_z^x \frac{d\zeta}{\lambda g(\zeta; q)}\right) dx \\ & \leq \lambda \|v - \hat{v}\| (1 + \theta) \|u\|_{\mathcal{H}} \|\mu_0\|. \end{aligned}$$

Combining these estimates, we find

$$\begin{aligned} \|Tv - T\hat{v}\|_{\mathcal{H}} & \leq \left( k_0 \lambda + \lambda^2 (1 + \theta P(\hat{v})) \|\mu_0\| + \lambda (1 + \theta) \|u\|_{\mathcal{H}} \|\mu_0\| \right) \|v - \hat{v}\|_{\mathcal{H}} \\ & = \lambda \left( k_0 + \|\mu_0\| (\lambda (1 + \theta P(\hat{v})) + (1 + \theta) \|u\|_{\mathcal{H}}) \right) \|v - \hat{v}\|_{\mathcal{H}}. \end{aligned}$$

Thus, we conclude that on every bounded subset  $M$  of  $H$  and for every  $u \in H$ , there exists  $\lambda_0$  such that  $T$  is a contraction in  $M$  for all  $\lambda \leq \lambda_0$ . Using



the arguments and results of Lemma 3.5, we have that if  $u \in C(r)$  and  $v \in C(r/(1 - \lambda_0 a))$ , then

$$\|Tv\|_{\mathcal{H}} \leq \lambda a \frac{r}{1 - \lambda_0 a} + r \leq \frac{r}{1 - \lambda_0 a}$$

for all  $\lambda \leq \lambda_0$ . Therefore, there exists a  $\lambda_0$ , such that for all  $\lambda \leq \lambda_0$ , the mapping  $T$  is a contraction on  $C(r/(1 - \lambda_0 a))$ .  $\square$

*Proof of Theorem 3.2.* Using Lemma 3.6, there exists a unique fixed point  $u_\lambda$  of  $T$  in  $H$ . It is not difficult to verify that  $u_\lambda \in \mathcal{D}(A)$  and  $u_\lambda$  is a solution of (a).  $\square$

We return to Theorem 3.1. Combining the results of subsections 3.1, 3.2 and 3.3, we find that the operator  $A$  has the following properties: (i)  $\mathcal{D}(A)$  is dense in  $H$ ; (ii)  $A - \lambda I$  is  $m$ -dissipative for all  $\lambda \geq k_0$ . Thus, by the Lumer-Phillips theorem (see [9]),  $A$  is the infinitesimal generator of a  $C_0$ -semigroup of bounded linear operators in  $H$  and we have established the results of Theorem 3.1.

#### 4. A nonlinear perturbation of the model

A nonlinear version of the model (1.1)–(1.2) in Section 1 is given by

$$(4.1) \quad \begin{aligned} \partial_t u(t, x, q) = & -\partial_x (g(x; q)u(t, x; q)) - \mu(x; q)u(t, x; q) \\ & - \mu_0(x; q)P(u(t))u(t, x; q), \quad 0 < x < x_{\max}(q), \end{aligned}$$

$$(4.2) \quad g(0; q)u(t, 0; q) = \int_Q \int_0^{x_{\max}(\hat{q})} K(q, \hat{q}, x)u(t, x; \hat{q}) dx dm(\hat{q}), \quad t > 0, q \in Q,$$

where  $P(u(t)) = \int_Q \int_0^{x_{\max}(\hat{q})} u(t, x; \hat{q}) dx dm(\hat{q})$ .

The new model can be viewed as a nonlinear perturbation of the linear model in the form

$$\frac{d}{dt} u(t) = (A + F)u(t)$$

where the operators  $A$  and  $F$  are defined in the previous section. Let  $p$  be the functional defined in Section 3 and the set  $C(r)$  is defined as before. It is not difficult to see that the nonlinear operator  $F + k_0 I : H \rightarrow H$  is locally Lipschitz continuous in the sense defined in [8] and the linear operator  $A - k_0 I$  generates a  $C_0$ -semigroup of contractions. We may conclude the following.

**Theorem 4.1.** *There exists a continuous nonlinear semigroup of operators  $\{S(t), t \geq 0\}$  on  $H$  satisfying*

- (a)  $S(t)u = e^{At}x + \int_0^t e^{A(t-\xi)}FS(\xi)u d\xi,$
- (b)  $\|S(t)u\|_{\mathcal{H}} \leq e^{at}\|u\|_{\mathcal{H}}$

for all  $u \in H$ . Moreover,  $A + F$  is the infinitesimal generator of the semigroup  $\{S(t), t \geq 0\}$ .

*Proof.* Using the results in Section 3 (more precisely Theorem 3.2), we may appeal to Theorems 1 and 2 of [8] with the convex functional  $p(u) = \|u\|_{\mathcal{H}}$ .  $\square$

## 5. Concluding remarks

In the above presentation, we have given a development of a linear semigroup formulation for rate distribution heterogeneous population models. This presentation is structured so that extension to a nonlinear version of the model in which the mortality rate depends on the total population is readily obtained. This is achieved by treating the nonlinear model as a perturbation of the linear model. If we include a nonlinearity in the birth process, the system can no longer be viewed as one defined by the perturbation of a linear operator by an additive locally Lipschitz continuous nonlinear perturbation. However, it may still be possible to adapt the ideas of [8] to obtain a nonlinear semigroup. This would require one to consider the two sources of nonlinearity separately. These and other related ideas are the subject of current efforts.

## References

1. H. T. BANKS, L. W. BOTSFORD, F. KAPPEL, C. WANG, *Modeling and estimation in size structured population models* Math. Ecology, T. G. Hallam, L. J. Gross, S. A. Levin (Eds.), Singapore: World Scientific, 1988, pp. 521-541.
2. H. T. BANKS and B. G. FITZPATRICK, *Estimation of growth rate distribution in size structure population models*, Quart. Appl. Math. **49** (1991), 215-235.
3. H. T. BANKS and F. KAPPEL, *Transformation semigroups and  $L^1$ -approximation for size structured population models*, Semigroup Forum **38** (1989), 141-155.
4. N. DUNFORD and J. T. SCHWARTZ, *Linear Operators*, Wiley-Interscience, New York, 1971.
5. B. G. FITZPATRICK, *Vector valued measure approach for a size structured population model*, J. Math. Anal. Appl. **172** (1993), 73-91.
6. W. HUYER, *A size structured population model with dispersion*, Institute für Mathematik Technische Universität Graz - Universität Graz Preprint No. 212-1992, May 1992; J. Math. Analysis Appl., to appear.
7. A. OKUBO, *Diffusion and Ecological Problems: Mathematical Models*, Springer-Verlag, Berlin, 1980.
8. S. OHARU and T. TAKAHASHI, *Locally Lipschitz continuous perturbations of linear dissipative operators and nonlinear semigroups*, Proceedings of the American Mathematical Society **100** (1987), 187-194.
9. A. PAZY, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer, New York, 1983.
10. G. F. WEBB, *Theory of nonlinear age-dependent population dynamics*, Marcel Dekker Inc., New York, 1985.

H. T. Banks  
Center for Research in Scientific Computation  
North Carolina State University  
Raleigh, NC 27695-8205, USA  
E-mail: htbanks@crsc1.math.ncsu.edu

F. Kappel  
Institut für Mathematik  
Universität Graz  
Heinrichstrasse 36, A8010 Graz, Austria  
E-mail: kappel@edvz.uni-graz.at

C. Wang  
Department of Mathematics  
University of Southern California  
Los Angeles, CA 90089-1113, USA  
E-mail: cwang@mathp.usc.edu

## DAMAGE DETECTION AND CHARACTERIZATION IN SMART MATERIAL STRUCTURES

H. T. BANKS AND Y. WANG

Center for Research in Scientific Computation  
North Carolina State University

**ABSTRACT.** We present theoretical, computational and experimental findings in initial investigations related to methods for detection and geometric characterization of damage in piezoceramic based smart material structures. The feasibility of using self-excitation/self-sensing with piezoceramics in vibration nondestructive testing is demonstrated using a combination of experimental and simulated data computational tests.

**1991 Mathematics Subject Classification.** 35R30, 73D50, 73K05

**Key words and phrases.** Inverse problems, distributed parameter systems, smart materials, damage detection.

### 1. Introduction

It has been known for some time that damage such as cracks, corruptions, and delaminations in a structure produces changes in mass, stiffness, damping and other characteristics and material parameters in dynamic response models for the structure. Our focus in this paper is the development of vibration response ideas for piezoceramic based smart material structures in which self-testing nondestructive evaluation (NDE) techniques may be employed. There are several ways to formulate and analyze concepts and questions related to this quest. One could use the co-called "method-of-maps" as done in the thermal based tomography techniques of [BK1, BK2, BKW] wherein the damaged physical domain (the damaged beam, plate, elastic structure) is mapped into a regular domain on which dynamic equations with damage dependent coefficients must hold. An alternative technique involves direct estimation of the irregular damaged domain (i.e., structures with holes, cracks, corruptions, etc.) on which the usual dynamic equations hold.

---

Research supported in part by the Air Force Office of Scientific Research under grants AFOSR F49620-93-1-0198 and AFOSR F49620-93-1-0280. Part of this research was carried out while both authors were visiting scientists at the Institute for Computer Application in Science and Engineering (ICASE), NASA Langley Research Center, which is supported by NASA under Contract Nos. NAS1-18605 and NAS1-19480.

The approach we take here is, strictly speaking, neither of the above, but it is more in the spirit of the latter one; we attempt to estimate damage-related changes in actual physical parameters (stiffness, damping) based on changes in physical geometry and characteristics.

Most of the previous efforts in the substantial literature on vibration related damage detection are based on modal methods (e.g. [ABB, AC, ACPS, CR, IC, S]). The basis for such methods is that damage produces a decrease in dynamic stiffness  $EI$ . This decrease in turn produces decreases in natural frequencies for an undamped simple beam (recall the eigenvalues are given by  $\lambda \sim \sqrt{EI/\rho A}$  where  $A$  is the cross sectional area of the beam and  $\rho$  is the mass density). While modal based methods may have certain advantages (e.g., they are simple if they do work, they can be used in development of rank-ordering of fractional eigenfrequency shift schemes that are reportedly insensitive to the magnitude of the damage [ABB]), modal based methods possess a number of major disadvantages. First of all, some of the modal based method investigations (e.g. [CR]) provide a strong argument for including geometry of the damage in any NDE testing scheme, something which is not easily done in frequency based methods. Indeed, mode and frequency characterizations are not so simple in variable structure systems: there is ample evidence (e.g. see [S]) that one should not use modal methods based on uniform undamped simple beams or plates as is often done in the engineering literature in addressing damage assessment methodologies. Since material parameters are most properly considered as spatially dependent quantities with damage manifested by changes in geometry (and hence in the spatial dependence of these parameters), it is unlikely that any rigorous theoretical basis for modal based methods for variable material structures will emerge. But perhaps the most serious objection to modal based methods resides in the fact that modal based methods have been shown to be highly unreliable (indeed inadequate! - see [BI] and the references therein) for estimation of variable material parameters such as damping in composite material structures.

In light of the above comments, a question of rather great interest then is: can one develop analytically sound, non-modal based self-excitation/self-sensing methods for detection and characterization (geometrical and quantitative) of damage in smart material structures? Here we address this question in the context of embedded piezoceramic structures.

For an embedded piezoceramics smart material damage detection and characterization methodology, there are several distinct requirements. These include:

- (a) One must be able to *estimate reliably* (repeatable across experiments) the *variable structure material parameters* of a piezoceramic loaded structure. This must be done using piezo actuation and sensing with accuracy comparable to that achievable with standard actuating (impulse hammers, solenoidal actuators) and sensing (accelerometers, strain gauges, laser vibrometers) devices in non-smart material testing schemes.
- (b) One must be able to use the actuation and sensing properties of the piezo-

ceramics to *excite the structure and analyze the response* (in a single experiment) for a reliable methodology that is the basis of self-excitation/self-sensing.

- (c) One must be able to *detect and characterize damage via vibration self-excitation/ self-sensing* that relies only on the input/output signals for the piezoceramics.

In this paper we address each of these requirements in the context of a piezo-ceramic loaded beam. This particular structure is sufficiently representative to make a compelling case for feasibility of the ideas we propose. In Section 2 below, we summarize a rigorous theoretical well-posedness and approximation foundation for the distributed parameter identification methodology that is the focus of our recent efforts. This is followed by Section 3 in which we detail results from experiments and computations that provide a favorable comparison for piezo actuating/sensing with standard methods (requirement (a) above). We then present experimental/computational findings on self-excitation/self-sensing capabilities of piezoceramic loaded beams. These findings suggest that requirement (b) above can indeed be satisfied.

Finally, in Section 4 we offer preliminary simulation findings on the use of vibration experiments to detect and characterize (geometrically) the extent of several types of damage in a beam using piezo actuation and sensing. Experimental verification of these positive findings with regard to requirement (c) is currently underway.

## 2. Distributed parameter identification methodology

The system we consider here is a cantilevered beam with piezoelectric ceramic patches for actuation and sensing. Our choice of structure is motivated by its simplicity and its representative nature. This configuration has also been well studied by conventional approaches such as finite element methods and provides a standard test bed model for comparison. The structure and model reveal the difficulties and possibilities inherent in developing models and methods for more complex structures containing damages.

We consider a cantilevered Euler-Bernoulli beam of length  $\ell$  fixed at  $x = 0$  and free at  $x = \ell$ . The transverse vibrations  $y = y(t, x)$  are described by the system

$$(2.1) \quad \begin{aligned} \rho(x) \frac{\partial^2 y}{\partial t^2}(t, x) + \gamma \frac{\partial y}{\partial t}(t, x) + \frac{\partial^2 M}{\partial x^2}(t, x) &= \tilde{f}(t, x), \quad 0 < x < \ell, \quad t > 0, \\ y(t, 0) = \frac{\partial y}{\partial x}(t, 0) &= 0, \quad M(t, \ell) = \frac{\partial M}{\partial x}(t, \ell) = 0, \end{aligned}$$

where  $\rho(x)$  is the linear mass density,  $\gamma$  is the coefficient of viscous (air) damping,  $M(t, x)$  is the internal moment and  $\tilde{f}(t, x)$  represents the external loads. For a simple Euler-Bernoulli beam with Kelvin-Voigt or strain rate damping, the internal

moment is composed of two components representing resistance to bending (with coefficient  $EI(x)$ ) and damping (with coefficient  $c_D I(x)$ ):

$$(2.2) \quad M(t, x) = EI(x) \frac{\partial^2 y}{\partial x^2}(t, x) + c_D I(x) \frac{\partial^3 y}{\partial x^2 \partial t}(t, x).$$

If piezoelectric elements are bonded to the beam in a configuration to produce (or sense) only bending, we have an actuator contribution  $M_p(t, x)$  in the form of an input moment (or voltage output proportional to the accumulated strain in the element). For a pair of piezoelectric actuators located between  $x_1$  and  $x_2$  on opposite sides of the beam excited by a voltage  $u(t)$  in an out-of-phase manner (see [BSW, CA, CFW, DFR, FC]), this moment term has the representation

$$(2.3) \quad M_p(t, x) = K_B \{H(x - x_1) - H(x - x_2)\} u(t)$$

where  $H(x)$  is the Heaviside or unit step function and  $K_B$  is a piezoelectric material parameter depending on the material piezoelectric properties as well as geometry. When the moment in (2.3) is added to that of (2.2) and substituted into (2.1), we obtain the model

$$(2.4) \quad \begin{aligned} \rho \frac{\partial^2 y}{\partial t^2} + \gamma \frac{\partial y}{\partial t} + \frac{\partial^2}{\partial x^2} \left( EI \frac{\partial^2 y}{\partial x^2} + c_D I \frac{\partial^3 y}{\partial x^2 \partial t} \right) \\ = K_B \left( \frac{d}{dx} \delta(x - x_2) - \frac{d}{dx} \delta(x - x_1) \right) u(t) + \tilde{f}(t, x) \\ y(t, 0) = \frac{\partial y}{\partial x}(t, 0) = 0, \quad M(t, l) = \frac{\partial M}{\partial x}(t, l) = 0 \end{aligned}$$

where  $\delta$  is the Dirac delta function. This is formally equivalent to the equation in weak or variational form (we replace partial derivatives in time by subscript  $t$  and space by superscript  $'$ )

$$(2.5) \quad \begin{aligned} \langle \rho y_{tt} + \gamma y_t, \phi \rangle + \langle EI y'' + c_D I y_t' + K_B(H_1 - H_2) u(t), \phi'' \rangle = \langle \tilde{f}(t, x), \phi \rangle \\ y(t, 0) = y'(t, 0) = 0, \end{aligned}$$

for sufficiently smooth functions  $\phi$  satisfying  $\phi(0) = \phi'(0) = 0$ . Here  $H_i$  is the shifted Heaviside function  $H_i(x) = H(x - x_i)$ ,  $i = 1, 2$  and  $\langle \cdot, \cdot \rangle$  is the usual  $L_2$  inner product.

For the same configuration, when the beam is under deformation (bending), the generated charges in terms of voltage across the piezoelectric sensors has the expression (see [DIC])

$$K_s \int_{x_1}^{x_2} \frac{\partial^2 y}{\partial x^2}(t, x) dx = K_s \left( \frac{\partial y}{\partial x}(t, x_2) - \frac{\partial y}{\partial x}(t, x_1) \right)$$

where  $K_s$  is a sensor constant which is also a piezoelectric material properties and geometry related quantity.

The system (2.4) is a formal representation of the dynamics of a damped beam with piezoelectric actuators. To develop computational techniques (e.g., finite elements) based on rigorous convergence arguments, it is necessary to first have a precise formulation of this system.

We start with an abstract formulation. Let  $V$  and  $H$  be complex Hilbert spaces satisfying  $V \hookrightarrow H = H^* \hookrightarrow V^*$  (see [W] for the construction of this Gelfand triple), where we denote their topological duals by  $V^*$  and  $H^*$ , respectively. Let  $Q$  be the admissible parameter set. The general second order system we consider here is given by

$$(2.6) \quad \begin{aligned} &\langle \ddot{y}(t), \psi \rangle_{V^*, V} + \sigma_1(q)(y(t), \psi) + \sigma_2(q)(\dot{y}(t), \psi) = \langle f(t), \psi \rangle_{V^*, V} \text{ for } \psi \in V, \\ &y(0) = y_0, \quad \dot{y}(0) = y_1. \end{aligned}$$

Here we use  $\langle \cdot, \cdot \rangle_{V^*, V}$  to denote the usual [W] duality product. The term  $\sigma_1(q)$  and  $\sigma_2(q)$  are parameter dependent sesquilinear forms on  $V$  satisfying  $V$ -ellipticity and  $V$ -continuity conditions. That is, we assume that  $\sigma_1(q)$  and  $\sigma_2(q)$  satisfy

$$(2.7) \quad \operatorname{Re} \sigma_i(q)(\phi, \phi) \geq k_i |\phi|_V^2,$$

$$(2.8) \quad |\sigma_i(\phi, \psi)| \leq c_i |\phi|_V \cdot |\psi|_V, \quad i = 1, 2$$

for  $k_i, c_i > 0$ ,  $\phi, \psi \in V$ . Under weak assumptions on  $f$ , the system (2.6) has a unique solution.

**Theorem 2.1.** *If the sesquilinear forms  $\sigma_1$  and  $\sigma_2$  satisfy conditions (2.7) and (2.8) with  $\sigma_1$  symmetric and  $f \in L_2((0, T), V^*)$ , then, for each  $w_0 = (y_0, y_1) \in \mathcal{H} = V \times H$ , the initial value problem (2.6) has a unique solution  $w(t) = (y(t), \dot{y}(t)) \in L_2((0, T), V \times V)$ . Moreover, this solution depends continuously on  $f$  and  $w_0$  in the sense that the mapping  $\{w_0, f\} \rightarrow w = (y, \dot{y})$  is continuous from  $\mathcal{H} \times L_2((0, T), V^*)$  to  $L_2((0, T), V \times V)$ .*

For a detailed proof of this theorem see [BIW].

Returning to our beam problem, we define the sesquilinear forms by

$$(2.9) \quad \sigma_1(q)(y, \phi) = \langle EI y'', \phi'' \rangle_{L_2}$$

$$(2.10) \quad \sigma_2(q)(\dot{y}, \phi) = \langle c_D I \dot{y}'', \phi'' \rangle_{L_2} + \langle \gamma \dot{y}, \phi \rangle_{L_2}$$

with the spaces defined by  $V = H_L^2(0, \ell) = \{\phi \in H^2(0, \ell) \mid \phi(0) = \phi'(\ell) = 0\}$  and  $H = L_2(0, \ell)$  with weighted inner product  $\langle \cdot, \cdot \rangle_H = \langle \rho \cdot, \cdot \rangle_{L_2}$ . We assume throughout that  $\rho \in L_\infty(0, \ell)$  with  $\rho(x) \geq \alpha$  for some  $\alpha > 0$ . The term  $f(t) = f(t, x)$  is given by

$$(2.11) \quad f(t, x) = \frac{1}{\rho} K_B (H''(x - x_2) - H''(x - x_1)) u(t) + \frac{1}{\rho} \tilde{f}(t, x), \quad 0 \leq x, x_1, x_2 \leq \ell,$$



where  $f(t, \cdot)$  belongs to the dual space (see [W])  $V^* = (H_L^2(0, \ell))^*$  (we assume that  $\hat{f}(t, \cdot) \in V^*$  which is a weak assumption on  $\hat{f}$ ). If both  $|EI(\cdot)|_{L_\infty} \geq \alpha$  and  $|c_D I(\cdot)|_{L_\infty} \geq \alpha$  for some  $\alpha > 0$ , then  $\sigma_1$  and  $\sigma_2$  are  $V$ -elliptic and continuous with  $\sigma_1$  symmetric; hence by Theorem 2.1 our beam equation is well posed for  $f(t, x)$  given by (2.11).

The parameter estimation problems can be stated in terms of finding parameters which give the best fit of the parameter dependent solutions of the partial differential equations to the observation data for response of the system to various excitations. In our case, the parameters to be estimated include beam mass density  $\rho(x)$ , stiffness coefficient  $EI(x)$  as well as damping parameters  $c_D I(x)$ ,  $\gamma$  and piezoelectric material parameters  $K_B, K_s$ . Let the collection of unknown parameters be denoted by  $q = (\rho(x), EI(x), c_D I(x), \gamma, K_B, K_s)$ . For given observations  $\{z_i\}$  corresponding to measurements at times  $t_i$  as obtained in most practical cases, we consider the least squares estimation problem of minimizing over  $q \in Q$  the least squares functional

$$(2.12) \quad J(y, z; q) = \left| \hat{C}_2 \left( \hat{C}_1 \{y(t_i, \cdot; q)\} - \{z_i\} \right) \right|^2,$$

where  $\{y(t_i, \cdot; q)\}$  are the parameter dependent weak solutions of (2.4) or (2.6) evaluated at each time  $t_i$ ,  $i = 1, 2, \dots, \bar{N}$  and  $|\cdot|$  is an appropriately chosen Euclidian norm. The set  $Q$  is some admissible parameter set. The operator  $\hat{C}_1$  may have several forms depending on the type of sensors being used. When the collected data are displacement, velocity, or acceleration at a point  $\bar{x}$  on the beam, we minimize

$$(2.13) \quad J_\nu(q) = \sum_{i=1}^N \left| \frac{\partial^\nu y}{\partial t^\nu}(t_i, \bar{x}; q) - z_i \right|^2,$$

for  $\nu = 0, 1, 2$ , respectively. In this case the operator  $\hat{C}_1$  involves differentiation (either  $\nu = 0, 1$  or  $2$  times, respectively) with respect to time followed by pointwise evaluation in  $t$  and  $x$ . When a piezoelectric sensor is used, the functional to be minimized is

$$(2.14) \quad J_p(q) = \sum_{i=1}^N \left| K_s \left( \frac{\partial y}{\partial x}(t_i, x_2; q) - \frac{\partial y}{\partial x}(t_i, x_1; q) \right) - z_i \right|^2,$$

for the piezoelectric elements being located on the beam between  $x_1$  and  $x_2$ . Here  $\{z_i\}$  are the measured voltages across the piezoelectric elements.

The operator  $\hat{C}_2$  may be the identity (corresponding to time domain identification procedures as in (2.13) and (2.14)) or the Fourier transform (corresponding to identification in the frequency domain). If the identification is carried out in the frequency domain and the operator  $\hat{C}_2$  is a Fourier transform operator, the

corresponding cost functional is given by

$$(2.15) \quad J(q) = \sum_{l=1}^{N_M} \left( \epsilon_1 \left| f_{k_l^y}(q) - f_{k_l^z} \right|^2 + \epsilon_2 \sum_{j=-n_l}^{N_l} \left| \|U(k_l^y + j; q)\| - \|Z(k_l^z + j)\| \right|^2 \right),$$

where  $U(k; q)$  and  $Z(k)$  are the Fourier series coefficients of  $\tilde{C}_1\{y(t_i, \bar{x}; q)\}$  and  $\{z_i\}$  respectively,  $f_{k_l^y}$  and  $f_{k_l^z}$  are the  $(k_l^y)^{th}$  vibration frequency of the solution  $U(k; q)$  and the  $(k_l^z)^{th}$  frequency of the observation data  $Z(k)$ ,  $\epsilon_1$ ,  $\epsilon_2$  are weighting constants, and  $n_l$ ,  $N_l$  are certain lower and upper limits associated with the width (or the support) of the  $l^{th}$  spikes. In obtaining (2.15), we have assumed that there are a finite and distinct number  $N_M$  ( $< \bar{N}$ ) of "spikes", i.e. vibration frequencies, among the  $Z(k)$  and the number of spikes of the solution  $U(k; q)$  is the same as  $N_M$ . We refer to [BW] for a detailed derivation and discussion of the cost functional (2.15).

The minimization in our parameter estimation problems involves an infinite dimensional state space and an infinite dimensional admissible parameter set (of functions). We thus consider Galerkin type approximations in the context of the variational formulation of (2.6). Let  $H^N$  be a sequence of finite dimensional subspaces of  $H$ , and  $Q^M$  be a sequence of finite dimensional sets approximating the parameter set  $Q$ . We define the orthogonal projections  $P^N : H \rightarrow H^N$  of  $H$  onto  $H^N$ . Then a family of approximating estimation problems with finite dimensional state spaces and parameter sets can be formulated by seeking  $q \in Q^M$  which minimizes

$$(2.16) \quad J^N(y^N, z; q) = \left| \tilde{C}_2 \left( \tilde{C}_1\{y^N(t_i, \bar{x}; q)\} - \{z_i\} \right) \right|^2,$$

where  $y^N(t; q) \in H^N$  is the solution to the finite dimensional approximation of (2.6) given by

$$(2.17) \quad \begin{aligned} & \langle \ddot{y}^N(t), \psi \rangle_{V^*, V} + \sigma_1(q) \langle y^N(t), \psi \rangle + \sigma_2(q) \langle \dot{y}^N(t), \psi \rangle = \langle f(t), \psi \rangle_{V^*, V} \\ & y^N(0) = P^N y_0, \quad \dot{y}^N(0) = P^N y_1, \end{aligned}$$

for  $\psi \in H^N$ .

For the parameter sets  $Q$  and  $Q^M$ , and state spaces  $H^N$ , we make the following hypotheses:

- (H1) The sets  $Q$  and  $Q^M$  lie in a metric space  $\tilde{Q}$  with metric  $d$  with  $Q$ ,  $Q^M$  compact in this metric and there is a mapping  $i^M : Q \rightarrow Q^M$  so that  $Q^M = i^M(Q)$ . Furthermore, for each  $q \in Q$ ,  $i^M(q) \rightarrow q$  in  $\tilde{Q}$  with the convergence uniform in  $q \in Q$ ;

- (H2) The finite dimensional subspaces  $H^N$  satisfy  $H^N \subset V$  as well as the approximation properties: For each  $\psi \in V$ ,  $\|\psi - P^N \psi\|_V \rightarrow 0$  as  $N \rightarrow \infty$ .

In addition to (uniform in  $Q$ ) ellipticity and continuity conditions (2.7) and (2.8), the sesquilinear forms  $\sigma_1(q)$  and  $\sigma_2(q)$  are assumed to be defined on  $Q$  and satisfy the continuity-with-respect-to-parameter conditions

$$(2.18) \quad |\sigma_i(q)(\phi, \psi) - \sigma_i(\tilde{q})(\phi, \psi)| \leq \gamma_i d(q, \tilde{q}) \|\phi\|_V \|\psi\|_V, \quad i = 1, 2,$$

for  $\phi, \psi \in V$  and  $q, \tilde{q} \in Q$  where the constants  $\gamma_1, \gamma_2$  depend only on  $Q$ .

Solving the approximate estimation problems involving (2.16) and (2.17), we obtain a sequence of estimates  $\{q^{N,M}\}$ . To obtain parameter estimate convergence and continuous dependence (with respect to the observations  $\{z_i\}$ ) results, it has been shown in [B], [BK3] and [BW1] that it suffices, under the assumption that  $Q$  is a compact set, to argue that for arbitrary  $\{q^{N,M}\} \in Q$  with  $q^{N,M} \rightarrow q \in Q$ , we have

$$(2.19) \quad \tilde{C}_2 \tilde{C}_1 y^N(t; q^{N,M}) \rightarrow \tilde{C}_2 \tilde{C}_1 y(t; q)$$

for each  $t$ .

The following theorem and corollary establish the convergence (2.19) for all forms of  $\tilde{C}_1, \tilde{C}_2$  that are of interest to us here.

**Theorem 2.2.** Suppose that  $Q, Q^M$  satisfy (H1) and  $H^N$  satisfies (H2). Assume that the sesquilinear forms  $\sigma_1(q)$  and  $\sigma_2(q)$  satisfy (2.7), (2.8), (2.18). Furthermore, assume that

$$q \mapsto f(t; q) \text{ is continuous from } Q \text{ to } L_2(0, T; V^*).$$

Let  $q^{N,M}$  be arbitrary in  $Q$  such that  $q^{N,M} \rightarrow q$  in  $Q$ ; then for  $t > 0$  as  $N, M \rightarrow \infty$  we have

$$\frac{\partial^\nu y^N}{\partial t^\nu}(t; q^{N,M}) \rightarrow \frac{\partial^\nu y}{\partial t^\nu}(t; q) \quad \text{in } V\text{-norm for } \nu = 0, 1, 2,$$

where  $y^N, \tilde{y}^N$  are the solutions to (2.17) and  $y, \tilde{y}$  are the solutions to (2.6).

Furthermore we have

**Corollary 2.3.** Under the assumptions of Theorem 2.2, we have as  $N, M \rightarrow \infty$

$$\frac{\partial y^N}{\partial x}(t; q^{N,M}) \rightarrow \frac{\partial y}{\partial x}(t; q) \quad \text{in } H^1\text{-norm, for } t > 0,$$

$$\sum_{i=1}^{N_M} \left( \epsilon_1 \left| f_{k_i^{y^N}}(q^{N,M}) - f_{k_i^y}(q) \right|^2 + \epsilon_2 \sum_{j=-n_i}^{N_i} \left| \|U^N(k_i^{y^N} + j; q^{N,M})\| - \|U(k_i^y + j; q)\| \right|^2 \right) \rightarrow 0,$$

where  $U^N(k; q^{N,M})$  and  $U(k; q)$  are the Fourier coefficients for  $\tilde{C}_1\{y^N(t; q^{N,M})\}$  and  $\tilde{C}_1\{y(t; q)\}$ , respectively.

The corollary follows from Theorem 2.2. For a detailed proof when  $\tilde{C}_2$  is the Fourier transform see [BWI]. For the proof of Theorem 2.2, we refer the reader to [BWIS] for the cases  $\nu = 0, 1$  and [BR] for the case  $\nu = 2$ .

We have outlined a framework for methods to carry out parameter identification for distributed parameter systems in both the time domain and the frequency domain. The motivation for introducing a frequency domain technique is that the time domain techniques give poor results when the observations contain several vibration modes and the initial guesses for parameters are not close to the optimal ones ([WA]). In such a situation we use a hybrid of these two methods: we iterate in a frequency domain formulation to obtain a reasonable initial parameter estimate for the time domain formulation which is then used to complete the identification procedure.

In Theorem 2.2, we presented convergence results allowing approximation of the parameter set. This form of the theoretical foundation provides a sound necessary basis for the detection and quantification of damage in structures since damage affects the dynamic behavior of systems in terms of changes in the system parameters such as  $\rho$  and  $I$ .

In concluding this section we make further comments related to the various allowable forms of the operator  $\tilde{C}_1$ . The flexibility in the choice of  $\tilde{C}_1$  and hence flexibility in the choice of sensors in the above theoretical framework makes it possible to carry out cross experiments to test the reliability of our methods. Even more important, this flexibility in our method enables us to use smart materials to carry out damage detection by employing only piezoelectric materials in actuating and sensing for the structure.

### 3. Quantitative and Experimental Results

In implementing the computational method outlined in Section 2, we chose cubic splines as the basis elements. The finite dimensional space  $H^N$  is defined as the span of the basis set for  $S_{3,B}^N(0, t)$  (the cubic B-splines corresponding to mesh size  $t/N$  - see [P]) with the basis set modified to satisfy the essential boundary conditions  $\phi(0) = \phi'(0) = 0$ . The finite dimensional space  $H^N$  constructed in this manner satisfies hypothesis (H2) (for details see [BWIS]). The dimension of the resulting approximation space is  $N + 1$ . For our calculations we chose  $N = 10$ , since the eigenvalues of this approximate finite dimensional system become stable at and after  $N = 10$  in the sense that the eigenvalues did not change significantly as  $N$  increases beyond 10.

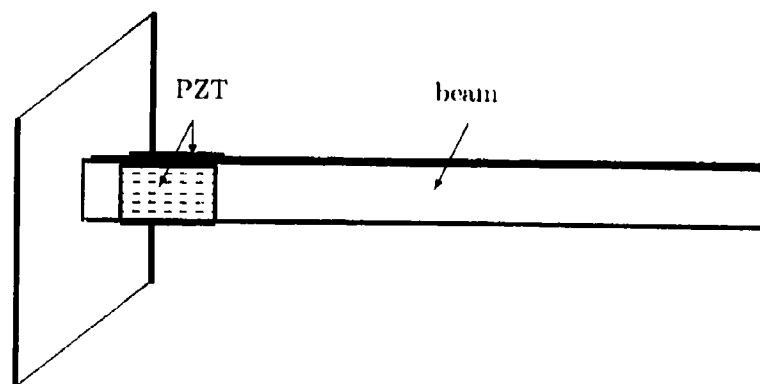
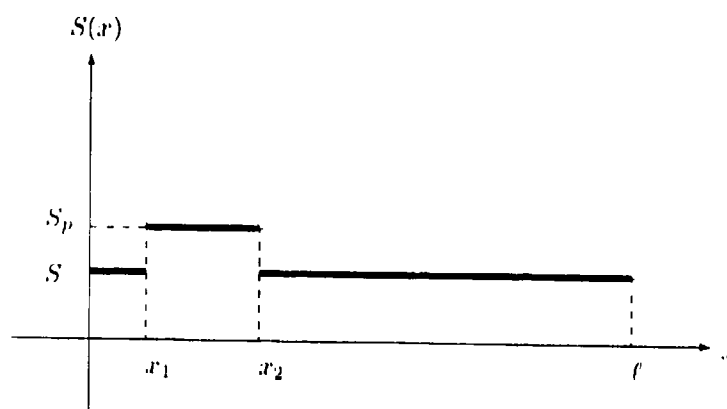


FIGURE 3.1. Test beam with piezoceramic patches.

FIGURE 3.2.  $\rho(x)$ ,  $EI(x)$ ,  $c_D I(x)$  function shape  $S(x)$ .

Unit	$\ell_b$	$w_b$	$t_b$	$\ell_p$	$w_p$	$t_p$
Metric	45.73cm	2.03cm	0.16cm	6.37cm	2.03cm	0.0254cm
English	18"	0.8"	.0625"	2.5"	0.8"	0.01"

TABLE 3.1. Dimension of the beam and piezoceramic patch.

Unit	$E_b$	$\rho_b$	$E_p$	$\rho_p$
Metric	$7.3 \times 10^6 N/cm^2$	$2.766g/cm^3$	$6.3 \times 10^6 N/cm^2$	$7.6g/cm^3$
English	$10.6 \times 10^6 psi$	$0.1lb/in^3$	$9.148 \times 10^6 psi$	$0.275lb/in^3$

TABLE 3.2. Characteristics of the beam and piezoceramic patch.

In the numerical examples we present next, an undamaged beam with two attached piezoceramic patches (PZT) as shown in Figure 3.1 was used as a test structure to demonstrate the consistency of identification results for our computational method across different excitation and sensing mechanisms. To accurately model the geometry of the structure, we assumed that the mass density, stiffness and Kelvin-Voigt damping coefficients are piecewise constant functions as depicted in Figure 3.2. These assumptions were subsequently verified to be most reasonable. For the choice of parameters  $q = (\rho(x), EI(x), c_D I(x), \gamma, K_B, K_s)$ , we take the parameter space  $\tilde{Q}$  to be  $[L_\infty(0, \ell)]^3 \times \mathbb{R}^3$ , and the parameter set  $Q$  to be a uniformly bounded collection of piecewise constant functions each having jump discontinuities at most at  $x_1$  and  $x_2$ . If we choose the mapping  $i^M$  as the identity, then hypothesis (H1) is satisfied.

Experiments with a test beam as depicted in Figure 3.1 were carried out at the Mechanical Systems Laboratory, then located at the State University of New York at Buffalo (now at Virginia Polytechnic Institute and State University). The measured dimensions of the actual structure are given in Table 3.1 and the handbook values for the physical characteristics of this aluminum (2024-T4) beam/piezo (G-1195 PZT ceramic) structure are given in Table 3.2. In the tables the subscripts indicate the materials:  $b$  for beam and  $p$  for piezoceramic,  $\ell$  is length,  $w$  is width,  $t$  is thickness,  $E$  is the Young's modulus, and  $\rho$  is the mass density. The patches were bonded to the beam on the opposite sides of the beam at the same position.

The beam was clamped at  $x = 0$ . The piezoceramic patches were placed between  $x_1 = 2.54cm$  and  $x_2 = 8.89cm$  (or 1" and 3.5" respectively) on the beam. The time response data and input signal from the experimental beam were obtained using a Tektronix Analyzer (model 2600). For more detailed descriptions of the experimental setup and the parameter identification procedure, see [BWIS].

In examining the ability of our method to estimate parameters reliably, independent of the type of excitation or sensing device used, two separate experiments

were performed. In the first experiment an impulse hammer was used as the input device and piezoceramic patches were used as a sensor. In the second experiment, the excitation was obtained via an input voltage to the patches and observation data was recorded from an accelerometer. At a sampling frequency of 256 Hz for 16 seconds, two modes (at 6.625 Hz and 38.375 Hz) were observed in the response in each experiment.

The results (graphs comparing the model response after ID of parameters with the experimental data) for the first identification experiment are given in Figure 3.3.

In this example, the beam was excited by an impulse force applied (via the impulse hammer) to the beam along the neutral axis at  $2.54\text{cm}(1'')$  away from the clamped end. Then the forcing function in (2.11) becomes (since  $u \equiv 0$ )

$$f(t, x) = \tilde{f}(t, x)$$

with  $\tilde{f}$  given by the impulse. The recorded signal exhibited a triangular shape. The voltages across the patches due to the beam vibrations were collected as observation data.

In the second experiment, a narrow triangular shaped voltages (approximating an impulse) was applied to the patches to excite the beam. The piezoceramic patches were excited out-of-phase so as to produce pure bending moments. Hence the input function is

$$f(t, x) = K_B(H''(x - x_2) - H''(x - x_1))u(t).$$

An accelerometer weighing 0.5 gram, located at  $\bar{x} = 2.14\text{cm}(0.844'')$  was used as the sensor. The parameter identification results are shown in Figure 3.4 where the data time history is compared to the model response corresponding to the best-fit parameters.

A summary of the estimated parameters is given in Table 3.3, where the units are for  $EI$  in  $N \cdot m^2$ ,  $\rho$  in  $kg/m$ ,  $c_D I$  in  $s \cdot N \cdot m^2$ ,  $\gamma$  in  $s \cdot N/m^2$ ,  $K_s$  in  $v$ , and  $K_B$  in  $N \cdot m/v$ . The measured and handbook quantities are listed in the table as "given" values; these were calculated according to the following equations

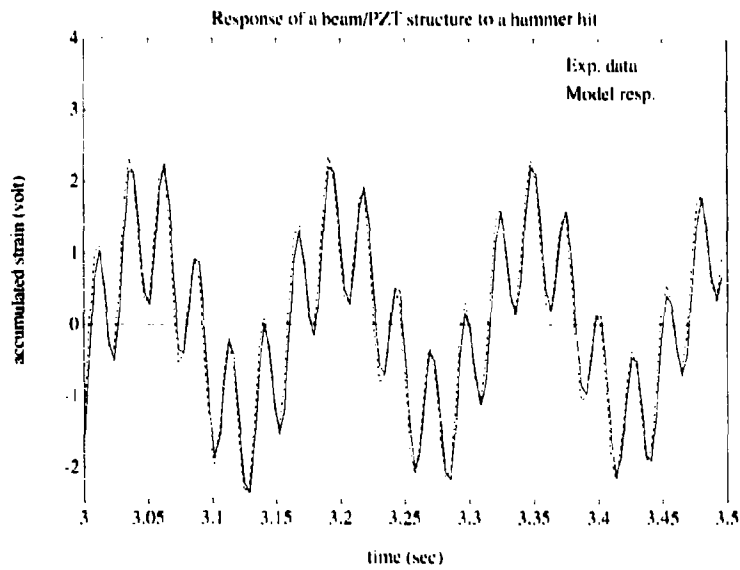
$$(3.1) \quad \rho(x) = \rho_b t_b w_b + 2\rho_p t_p w_p \chi_p(x)$$

$$(3.2) \quad EI(x) = \frac{1}{12} t_b^3 w_b E_b + \frac{2}{3} \left[ \left( \frac{t_b}{2} + t_p \right)^3 - \left( \frac{t_b}{2} \right)^3 \right] w_p E_p \chi_p(x)$$

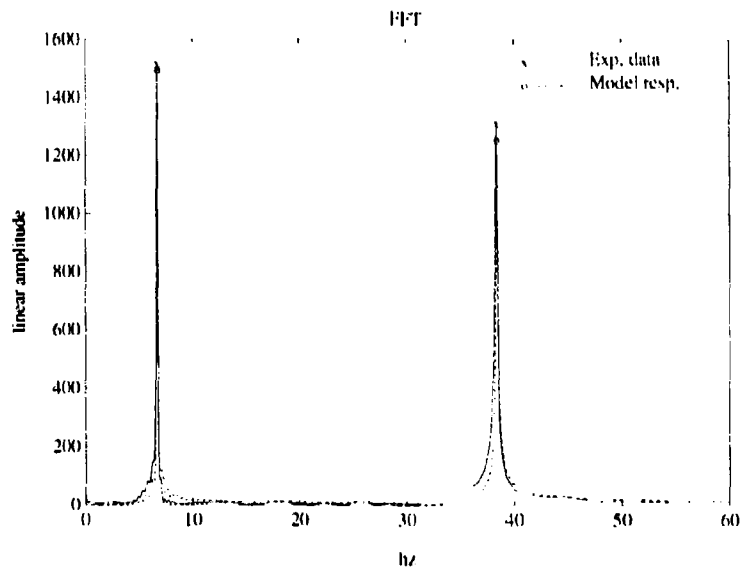
where  $\chi_p$  is a characteristic function given by

$$(3.3) \quad \chi_p(x) = \begin{cases} 1 & x_1 < x < x_2 \\ 0 & \text{otherwise.} \end{cases}$$

We next sought to demonstrate the capability of piezoelectric materials in smart structures (specifically in the area of self-evaluation of structures). We designed and performed an experiment on the same beam in which the piezoceramic



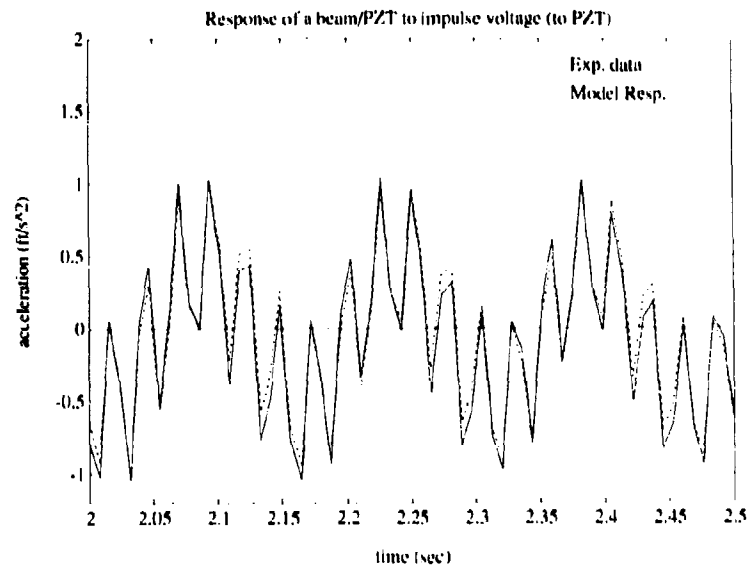
(a): Time history in [3, 3.5] seconds.



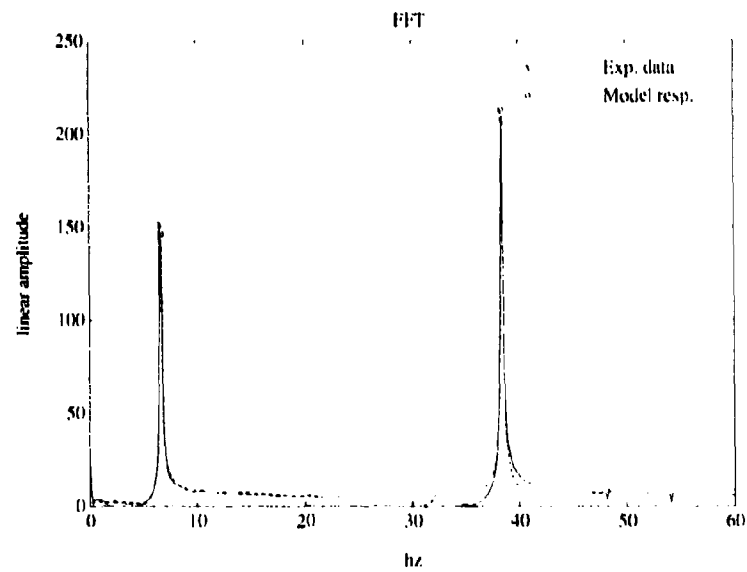
(b): Frequency content.

FIGURE 3.3. ID result for Experiment 1 (actuator impulse hammer, sensor PZT).





(a): Time history in [3, 3.5] seconds.



(b): Frequency content.

FIGURE 3.4. ID result for Experiment 2 (actuator PZT; sensor accelerometer).

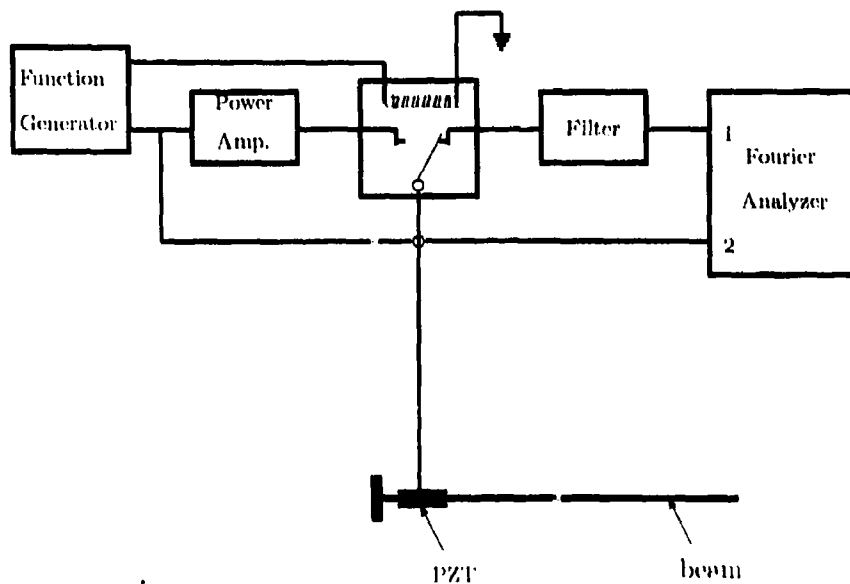


FIGURE 3.5. Circuit for the experiment using PZT as actuator/sensor.

patches were used as actuator and sensor in self-identification of material parameters. In this third experiment, the structure was excited by an input of sinusoidal voltage, out-of-phase, to the PZT patches for a time period of  $[0, t_0]$  and was left to free vibrate for  $t \geq t_0$ . The circuit setup for this experiment is depicted in Figure 3.5.

The accumulated strain in form of a voltage across the patches was recorded for approximately 14 seconds. The sample rate was 256 Hz and the excitation frequency was 38.25 Hz; this is approximately the second vibration mode of the composite structure (beam with bonded patches). The estimated parameters are reported in Table 3.4 in which the units are the same as those given for Table 3.3. Comparison of the model response (after parameter ID) to the experimental data is shown in Figure 3.6.

It can be observed that identification results for  $\rho(x)$ ,  $x_1 < x < x_2$ , in all three examples yield values that are substantially larger than the given one. One explanation for this is that equation (3.1) is simply a superposition of linear mass densities of the beam and the patches which ignores the mass of the bonding glue and the conducting copper foil.

Exp. No.	actuator / sensor	$EI$		$\rho$		$c_{PI} (\times 10^{-5})$		$\gamma$	$K_s$	$K_B$
		beam	beam + PZT	beam	beam + PZT	beam	beam + PZT			
given		0.495	1.051	0.089	0.168	—	—	—	—	—
Exp. 1	hammer / PZT	0.491	0.793	0.093	0.433	0.649	1.255	0.013	46.823	—
Exp. 2	PZT / accel.	0.505	0.798	0.096	0.441	0.637	1.275	0.013	—	0.0175

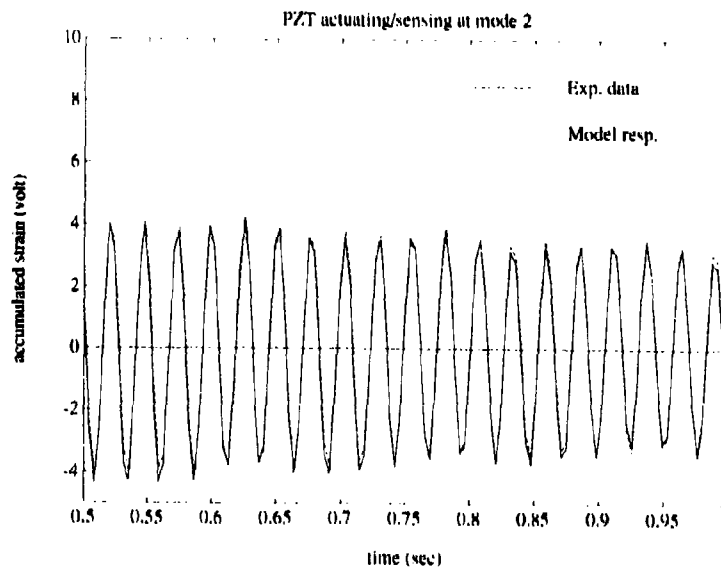
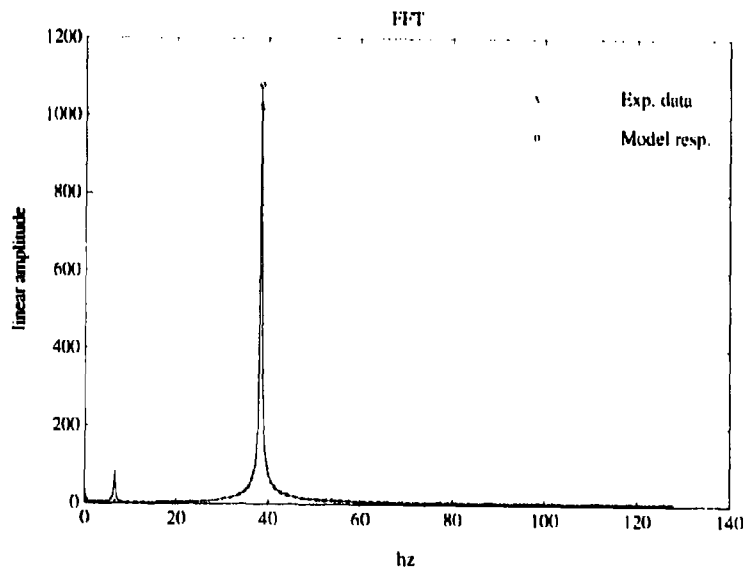
TABLE 3.3. Given and estimated structural parameters for Experiments 1 and 2.

Exp. No.	actuator / sensor	$EI$		$\rho$		$c_{PI} (\times 10^{-5})$		$\gamma$	$K_s$	$K_B$
		beam	beam + PZT	beam	beam + PZT	beam	beam + PZT			
given		0.495	1.051	0.089	0.168	—	—	—	—	—
Exp. 3	PZT / PZT	0.505	0.796	0.095	0.483	0.607	1.275	0.013	46.823	0.0196

TABLE 3.4. Given and estimated structural parameters for Experiment 3.

In all our parameter estimation procedures, we adopted a so-called hybrid method (mentioned previously), wherein we started with the cost function (2.15) (the cost function corresponding to frequency domain data) and then switched to (2.13) or (2.14) (time domain data). Our experience on these and numerous other inverse problems reveal that parameter identification in time domain is more sensitive and accurate if the initial guesses of parameters are close to the optimal ones. On the other hand, parameter identification in the frequency domain will yield quick and rough estimates with the resulting parameters in a neighborhood of the optimal values.

From the above reported results, we can conclude that the methodology outlined in the previous section yields theoretical, computational and experimental findings that are consistent. Moreover, it is feasible to use smart materials in non-destructive self-evaluation schemes such as presented in the next section.

(a): Time history in  $[0.5, 1.0]$  seconds.

(b): Frequency content.

FIGURE 3.6. ID result for Experiment 3 (actuator-PZT; sensor-PZT).

#### 4. Detection and Characterization of Damage

In our preliminary investigation of damage detection capabilities we have carried out simulation studies. The structure used in the simulation has the same dimensions and physical characteristics as the beam in Section 3. The difference is that the beam is assumed damaged by a hole. Since the beam model is based on Euler-Bernoulli theory, we assume that the center of the hole coincides with the neutral axis of the beam. Furthermore, we assume that the hole is either a circle centered at  $x_d$  with dimension  $r$  or a rectangle centered at  $x_d$  with dimension of  $2r \times w_d$ . With these assumptions, the functions  $\rho(x)$  and  $EI(x)$  can be expressed by

$$(4.1) \quad \rho(x) = \rho_b t_b w_b + 2\rho_p t_p w_p \chi_p(x) - \rho_b t_b (w_b - \tilde{w}) \chi_d(x)$$

$$(4.2) \quad EI(x) =$$

$$\frac{1}{12} t_b^3 w_b E_b + \frac{2}{3} \left[ \left( \frac{t_b}{2} + t_p \right)^3 - \left( \frac{t_b}{2} \right)^3 \right] w_p E_p \chi_p(x) - \frac{1}{12} t_b^3 E_b (w_b - \tilde{w}) \chi_d(x)$$

where the characteristic function  $\chi_p$  is the same as (3.3) and  $\chi_d$  is given by

$$(4.3) \quad \chi_d(x) = \begin{cases} 1 & x_{d1} < x < x_{d2} \\ 0 & \text{otherwise.} \end{cases}$$

Here  $x_{d1} = x_d - r$  and  $x_{d2} = x_d + r$ , and

$$\tilde{w} = \begin{cases} w_d & \text{for a rectangular hole,} \\ 2\sqrt{r^2 - (x - x_d)^2} & \text{for a circular hole.} \end{cases}$$

The Kelvin-Voigt damping function  $c_D I(x)$  has the same form as (4.2) with  $E$  replaced by  $c_D$ . In both examples, the damage is parameterized by parameters  $r$  and  $x_d$  with  $w_d$  held fixed in the case of a rectangular hole at a value  $w_d = 1.016\text{cm}(.4'')$ .

In the simulations, we assumed that we knew the shape (rectangular or circular) of the damage. The damage detection (location and size of the hole) is carried out through parameter identification by fitting the solution of the model to the simulated observation data much in the same manner as described for the experimental data cases described in Section 3. Voltage measurements across the piezoceramic patches are used as observations.

The input data is a triangular (time function) voltage in a very short time period (see Figure 4.1) to simulate an impulse input. The choice of the triangular function was motivated by impulse hammer excitation records from actual experiments as reported in Section 3.

The simulated "observation data"  $\{z_i\}$  were generated by solving the finite dimensional model (2.17). The number of finite elements was set at  $N = 10$  with cubic spline basis elements. The time interval was  $[0, 8]$  seconds with sampling rate

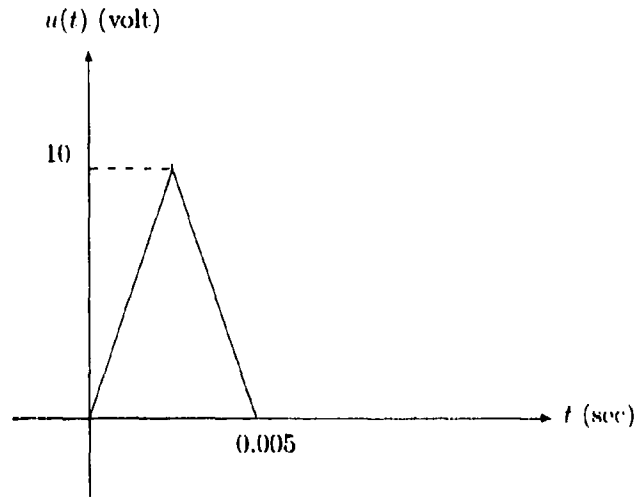


FIGURE 4.1. Input voltage.

512 Hz. Parameters estimated from experimental data and given in Table 3.4 were used for the simulated test structure.

In the following two examples, one with a rectangular hole and one with a circular hole, two cases were investigated. In Case I, the simulated observation data was generated with a damaged beam (a hole in the beam) and in Case II the data was generated with a undamaged beam (without holes). In these simple preliminary tests, these simulated data sets were used without additional noise. In the parameter estimation procedures, we also used 10 finite elements (cubic splines) in solving the inverse problems.

**EXAMPLE 1.** We assumed there was a rectangular hole in the beam. The parameters defining the hole in the simulated data were

$$r = 0.508cm(0.2''), \quad x_d = 10.414cm(4.1'').$$

The initial guesses and converged parameters are shown in Table 4.1.

**EXAMPLE 2.** In this example, the damage is characterized by a circular hole in the beam. The parameters for the hole were

$$r = 0.508cm(0.2''), \quad x_d = 10.414cm(4.1'').$$

The initial guesses and converged parameters are shown in Table 4.2.

As was noted, in these simulation examples we used noise free simulated observation data in the parameter estimation problems. We are currently studying the effects of noise in the data on the accuracy of the estimated parameters.

Case I.

Unit	Parameters	Initial	Final	Preassigned
Metric	$x_d(cm)$	3.048	10.414	10.414
	$r(cm)$	0.152	0.508	0.508
English	$x_d(in)$	1.2	4.1	4.1
	$r(in)$	0.06	0.2	0.2

Case II.

Unit	Parameters	Initial	Final	Preassigned
Metric	$x_d(cm)$	3.048	1.753	0
	$r(cm)$	0.152	$1.592 \times 10^{-6}$	0
English	$x_d(in)$	1.2	0.69	0
	$r(in)$	0.06	$6.27 \times 10^{-7}$	0

Case I. damage in shape of a rectangle; Case II. no damage

TABLE 4.1. Damage detection results

Case I.

Unit	Parameters	Initial	Final	Preassigned
Metric	$x_d(cm)$	3.048	10.414	10.414
	$r(cm)$	0.152	0.508	0.508
English	$x_d(in)$	1.2	4.1	4.1
	$r(in)$	0.06	0.2	0.2

Case II.

Unit	Parameters	Initial	Final	Preassigned
Metric	$x_d(cm)$	3.048	3.315	0
	$r(cm)$	0.152	$4.55 \times 10^{-3}$	0
English	$x_d(in)$	1.2	1.305	0
	$r(in)$	0.06	$1.791 \times 10^{-3}$	0

Case I. damage in shape of a circle; Case II. no damage

TABLE 4.2. Damage detection results

In many simulation studies found in the literature, the impulse function inputs are approximated by a smooth function, for example,  $\sin^2(\alpha t)$ . Our choice of a triangular shape for these input functions is in practice equivalent to the choice of a smooth function since discrete input data are used and hence smoothness of the time records does not play a role in the sampling process or the trajectories obtained.

The results reported in this section are very encouraging. These initial calculations suggest that our methods for parameter identification can reveal whether there is a damage or not. Our current efforts involve eliminating the assumed pre-knowledge of the shape of the damage. We also are investigating issues related to the detection and characterization of cracks in structures.

## 5. Concluding remarks

As we have noted in the introduction of this paper, the idea of using vibration testing as a basis for damage detection in structures is not new. However, most methods to date are based on modal techniques. In this paper we have presented a theoretical and computational non-modal framework for the identification of spatially dependent dynamic parameters in piezoceramic embedded structures using nondestructive vibration tests. This rigorous foundation permits one to use the piezoceramics to both excite and sense vibrations in a self-analysis framework that is a natural major feature of smart material structures.

Using data from beam experiments, we have demonstrated the feasibility of our approach in obtaining reliable physically meaningful dynamic parameters such as stiffness, damping, and mass density. In preliminary simulation studies on the possibility of detecting and geometrically characterizing damage such as holes, we have presented examples which suggest that such methods offer great promise. While much remains to be done in developing these ideas into general algorithms, we are most encouraged by our initial findings and are currently designing and carrying out physical experiments to support further efforts on geometric characterization of damage such as corosions, cracks, and delaminations in composite material structures.

## Acknowledgement

The authors would like to gratefully acknowledge continuing valuable discussions and encouragement from Dr. W. Winfree, NASA Langley Research Center, and Dr. D.J. Inman, Virginia Polytechnic Institute and State University. The authors are indebted to Dr. Inman and Dr. J. Slater (Wright State University) for their assistance and collaboration on the experimental efforts summarized in this paper and the references.



## References

- [ABB] D. Armon, Y. Ben-Haim, and S. Braum, *Crack detection in beams by rank-ordering of eigenfrequency shifts*, Mechanical Systems and Signal Processing, to appear.
- [AC] R.D. Adams and P. Cawley, *The location of defects in structures from measurements of natural frequencies*, J. Strain Analysis 14 (1979), pp. 49-57.
- [ACPS] R.D. Adams, P. Cauley, C.J. Pye and B.J. Stone, *A vibrational technique for non-destructively assessing the integrity of structures*, J. Mech. Engr. Sci. 20 (1979), pp. 93-100.
- [B] H.T. Banks, *On a variational approach to some parameter estimation problems*, in Distributed Parameter Systems (ed. by F. Kappel, et. al.), Springer Lecture Notes in Control and Information Sciences, 75 (1985), pp. 1-23.
- [BI] H.T. Banks and D.J. Inman, *On damping mechanisms in beams*, ICASE Rep. #89-64 August, 1989; ASME J. Applied Mechanics 58 (1991), pp. 716-723.
- [BIW] H.T. Banks, K. Ito and Y. Wang, *Well posedness for damped second order systems with unbounded input operators*, CRSC-TR93-10, North Carolina State University, June, 1993; Differential and Integral Equations, to appear.
- [BK1] H.T. Banks and F. Kojima, *Boundary shape identification problems in two dimensional domains related to thermal testing of materials*, Quart. Applied Math., 47 (1989), pp. 273-293.
- [BK2] H.T. Banks and F. Kojima, *Boundary identification for 2-D parabolic systems arising in thermal testing materials*, Proceedings of 27th IEEE Cong. on Dec. and Control (1988), pp.1678-1683.
- [BK3] H.T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhäuser, Boston, 1989.
- [BKW] H.T. Banks, F. Kojima, and W.P. Winfree, *Boundary estimation problems arising in thermal tomography*, CAMS Tech. Rep 89-6, University of Southern California; Inverse Problems 6 (1990), pp. 897-921.
- [BR] H.T. Banks and D.A. Rebnord, *Analytic semigroups; applications to inverse problems for flexible structures*, in Differential Equations with Applications (ed. by J. Goldstein, et. al.), Marcel Dekker, 1991, pp. 21-35.
- [BSW] H.T. Banks, R. Smith and Y. Wang, *The modeling of piezoceramic patch interactions with shell, plates and beams*, Quarterly of Applied Mathematics, to appear.
- [BW] H.T. Banks and Y. Wang, *Parameter identification in the frequency domain*, in Progress in Systems and Control Theory: Computation and Control III, (ed. by K Bowers and J. Lund) Birkhäuser, 1993, pp. 49-62.
- [BW1] H.T. Banks, Y. Wang and D.J. Inman, *Bending and shear damping in beams: frequency domain estimation techniques*, CAMS Tech. Rep. 91-25, September, 1991, USC; ASME J. Vibration and Acoustics, to appear.
- [BWIS] H.T. Banks, Y. Wang, D.J. Inman and J.C. Slater, *Approximation and parameter identification for damped second order systems with unbounded input operators*, CRSC-TR93-9, North Carolina State University, May, 1993; Control: Theory and Advanced Technology, to appear.
- [CA] E.F. Crawley and E.H. Anderson, *Detailed models for piezoceramic actuation of beams*, AIAA Conf. paper 89-1388-CP, 1989, pp. 2000-2010.
- [CFW] R.L. Clark, Jr., C.R. Fuller and A. Wicks, *Characterization of multiple piezoelectric actuators for structural excitation*, J. Acoust. Soc. Amer., 1991, to appear.
- [CR] P. Cawley and R. Ray, *A comparison of the natural frequency changes produced by cracks and slots*, ASME J. Vibration, Acoustics, Stress and Reliability in Design 110 (1988), pp. 366-370.
- [DFR] E.K. Dimitriadis, C.R. Fuller and C.A. Rogers, *Piezoelectric actuators for distributed noise and vibration excitation of thin plates*, Proc. 8th ASME Conf. on Failure Prevention, Reliability, and Stress Analysis, Montreal, 1989, pp. 223-233.

- [DIG] J. Dosch, D.J. Inman, and E. Garcia, *A self-sensing piezoelectric actuator for collocated control*, J. Intell. Material Systems and Structures, **3** 1992, pp. 166-185.
- [FC] J.L. Fanson and T.K. Caughey, *Positive position feedback control for large structures*, Structural Dynamics and Materials Conf., AIAA Conf. Paper 87-0902, 1987, pp. 588-598.
- [IC] A. Isman and K. Craig, *Damage detection in composite structures using piezoelectric materials*, 1st ARO Workshop on Smart Structures and Materials, U. Texas-Arlington, Sept. 22-24, 1993.
- [P] P.M. Prenter, *Splines and Variational Methods*, Wiley, New York, 1975.
- [S] H. Sato, *Free vibration of beams with abrupt changes of cross-section*, J. Sound and Vibration **89** (1983), pp. 59-64.
- [WA] Y. Wang, *Damping Modeling and Parameter Estimation in Timoshenko Beams*, Ph.D. Thesis, Brown University, May 1991.
- [W] J. Wloka, *Partial Differential Equations*, Cambridge University Press, New York, 1987.

H. T. Banks  
Center for Research in Scientific Computation  
North Carolina State University  
Raleigh, NC 27695-8205, USA  
E-mail: htbanks@crsc1.math.ncsu.edu

Y. Wang  
Center for Research in Scientific Computation  
North Carolina State University  
Raleigh, NC 27695-8205, USA  
E-mail: ywang@crsc1.math.ncsu.edu

## OPTIMALITY CONDITIONS FOR NON-QUALIFIED PARABOLIC CONTROL PROBLEMS

M. BERGOUNIOUX, T. MÄNNIKÖ, AND D. TIBA

Département de Mathématiques et d'Informatique  
U.F.R. Sciences Université d'Orléans

Department of Mathematics  
University of Jyväskylä

Institute of Mathematics of the Romanian Academy

**ABSTRACT.** We consider parabolic state constrained optimal control problems where the usual Slater condition is not necessarily satisfied. Instead, a weaker interiority property is assumed. Optimality conditions with a Lagrange multiplier are given. As an application we present an augmented Lagrangian algorithm. Numerical test results are included.

1991 *Mathematics Subject Classification.* 49K20, 49M29

*Key words and phrases.* Optimal control, optimality conditions, Lagrange multipliers, parabolic systems.

### 1. Introduction

Let us consider the following control problem with constraints:

$$(\mathcal{P}) \quad \text{Minimize } \left\{ \frac{1}{2} \int_Q (y - z_d)^2 dx dt + \frac{N}{2} \int_Q u^2 dx dt \right\}$$

subject to

$$(1.1) \quad \frac{\partial y}{\partial t} - \Delta y = f + u \quad \text{in } Q,$$

$$(1.2) \quad y(x, t) = 0 \quad \text{on } \Sigma,$$

$$(1.3) \quad y(x, 0) = y_0 \quad \text{in } \Omega,$$

$$(1.4) \quad u \in U_{ad},$$

$$(1.5) \quad y \in K_0,$$

where  $\Omega \subset \mathbb{R}^n$  is a bounded regular domain,  $Q = ]0, T[ \times \Omega$ ,  $T > 0$ ,  $\Sigma = [0, T] \times \partial\Omega$  and  $N > 0$ . Moreover,  $z_d, f \in L^2(Q)$ ,  $y_0 \in H_0^1(\Omega) \cap H^2(\Omega)$ , and  $U_{ad}, K_0$  are closed convex subsets of  $L^2(Q)$ .

We shall examine the cases where both  $U_{ad}$  and  $K_0$  have weak interiority properties of Slater type, or have even void interior in  $L^\infty(Q)$ . We give optimality conditions with a Lagrange multiplier which may be a function or a measure and which appears in a symmetric form with respect to the two constraints. As an application we present an augmented Lagrangian algorithm for the solution of  $(P)$ , and two numerical examples are included.

The approach of this paper is based on the penalization of the state system rather than of the constraints. An essential role is played by new estimates of the approximate Lagrange multiplier which can be mainly compared with the earlier works [1], [2], [3] of the first author. Another approach to such problems can be found in Bonnans and Casas [5], [6].

The main novelty of this paper consists in the treatment of the state constraints when  $K_0$  has a void interior even in  $L^\infty(Q)$ , that is without satisfying the Slater condition. It is shown that a weak interiority property of  $U_{ad}$  is sufficient in order to ensure the existence of the Lagrange multiplier.

## 2. Optimality conditions

In what follows, we consider the case where the constraint (1.5) is of the form

$$(2.1) \quad \varphi(x, t) \leq y(x, t) \leq \psi(x, t) \quad \text{in } Q.$$

We assume that  $\varphi, \psi \in C^1(Q)$ ,  $\varphi \leq \psi$  in  $Q$ ,  $\varphi \leq 0 \leq \psi$  on  $\Sigma$  and  $\varphi(x, 0) \leq y_0(x) \leq \psi(x, 0)$ . More specific conditions on  $\varphi, \psi$  and  $U_{ad}$  with examples will be given in §4.

Let us define

$$(2.2) \quad \mathcal{K} = \{y \in H^{2,1}(Q) \mid \varphi(x, t) \leq y(x, t) \leq \psi(x, t) \text{ a.e. in } Q, \\ y(x, 0) = y_0(x) \text{ a.e. in } \Omega\}.$$

We approximate  $(P)$  by the penalized, "decoupled" optimization problem:

$$(P_\varepsilon) \quad \begin{aligned} \text{Minimize } & \left\{ \frac{1}{2} \int_Q (y - z_d)^2 dx dt + \frac{N}{2} \int_Q u^2 dx dt \right. \\ & \left. + \frac{1}{2\varepsilon} \int_Q \left( \frac{\partial y}{\partial t} - \Delta y - f - u \right)^2 dx dt \right\} \end{aligned}$$

subject to  $u \in U_{ad}, y \in K$ .

We denote by  $[y^*, u^*]$  the unique optimal pair for  $(P)$  (which exists under the standard admissibility condition) and by  $[y_\varepsilon, u_\varepsilon]$  the optimal pair for  $(P_\varepsilon)$ . We

also denote

$$(2.3) \quad q_\varepsilon = \frac{1}{\varepsilon} \left( \frac{\partial y_\varepsilon}{\partial t} - \Delta y_\varepsilon - f - u_\varepsilon \right).$$

Moreover, let  $p_\varepsilon$  be the solution of the adjoint equation

$$(2.4) \quad -\frac{\partial p_\varepsilon}{\partial t} - \Delta p_\varepsilon = y_\varepsilon - z_d \quad \text{in } Q,$$

$$(2.5) \quad p_\varepsilon(x, t) = 0 \quad \text{on } \Sigma,$$

$$(2.6) \quad p_\varepsilon(x, T) = 0 \quad \text{in } \Omega,$$

and  $p^*$  the solution of (2.4)–(2.6) with  $y_\varepsilon$  replaced by  $y^*$ .

We may easily prove the following propositions (see Bergounioux [2], [3]):

**Proposition 2.1.** *When  $\varepsilon \rightarrow 0$ , we have*

$$(2.7) \quad y_\varepsilon \rightarrow y^* \quad \text{strongly in } H^{2,1}(Q),$$

$$(2.8) \quad u_\varepsilon \rightarrow u^* \quad \text{strongly in } L^2(Q),$$

$$(2.9) \quad p_\varepsilon \rightarrow p^* \quad \text{strongly in } H^{2,1}(Q),$$

$$(2.10) \quad \{\sqrt{\varepsilon} q_\varepsilon\} \quad \text{bounded in } L^2(Q).$$

**Proposition 2.2.** *The approximate optimality conditions may be written in the decoupled form*

$$(2.11) \quad \forall y \in K \quad \int_Q (p_\varepsilon + q_\varepsilon) \left( \frac{\partial}{\partial t} (y - y_\varepsilon) - \Delta (y - y_\varepsilon) \right) dx dt \geq 0,$$

$$(2.12) \quad \forall u \in U_{ad} \quad \int_Q (N u_\varepsilon - q_\varepsilon) (u - u_\varepsilon) dx dt \geq 0.$$

In order to obtain a good estimate also on  $\{q_\varepsilon\}$  and to pass to the limit in (2.11), (2.12), we study several cases of interest.

### Example 1

Let  $\varphi \equiv a$ ,  $\psi \equiv b$  in  $Q$  and  $U_{ad} = \{u \in L^2(Q) \mid c \leq u(x, t) \leq d \text{ a.e. in } Q\}$ . We strengthen the admissibility condition to the following Slater type assumption:

$$(2.13) \quad \exists \delta > 0, \exists \tilde{u} \in U_{ad} \quad \text{such that} \\ \tilde{y} = \mathcal{T}(\tilde{u}) \quad \text{satisfies} \quad a + \delta \leq \tilde{y}(x, t) \leq b - \delta \quad \text{in } Q,$$

where  $\mathcal{T}$  is the affine bounded operator  $u \mapsto y$  defined by (1.1)–(1.3). We notice that  $U_{ad} \subset L^\infty(Q)$  in this example, and the relations (2.7)–(2.9) are valid in  $L^p(Q)$ ,  $W^{2,1,p}(Q)$  for any  $p > 1$ .

Let us consider the pair  $[y_\kappa, \tilde{u}] = [\mathcal{T}(\tilde{u} - \rho\kappa), \tilde{u}]$ , where  $\tilde{u}$  is given by (2.13),  $\rho > 0$  will be precised later and  $\kappa \in L^p(Q)$ ,  $|\kappa|_{L^p(Q)} = 1$  is arbitrary. By the continuity with respect to the data of  $\mathcal{T}$ , we get

$$|\tilde{y} - y_\kappa|_{W^{2,1,p}(Q)} \leq C\rho$$

with  $C$  depending only on  $p$  and  $Q$ . If  $p > (n+2)/2$ , the anisotropic Sobolev embedding theorem gives

$$|\tilde{y} - y_\kappa|_{C(Q)} \leq C_1 C \rho$$

with  $C_1$  independent of  $\kappa, \rho$ . Taking  $\rho = \delta/(2C_1 C)$ , (2.13) gives  $y_\kappa \in K$  for all  $\kappa$ .

**Proposition 2.3.** *Let  $p$  and  $\rho$  be as above. Then  $\{q_\varepsilon\}$  is bounded in  $L^{p'}(Q)$ , where  $1/p + 1/p' = 1$ .*

**Proof.** We use the pair  $[y_\kappa, \tilde{u}]$  in (2.11), (2.12) and we add the two relations to obtain

$$\int_Q (p_\varepsilon + q_\varepsilon)(\tilde{u} - \rho\kappa - \varepsilon q_\varepsilon - u_\varepsilon) dx dt + \int_Q (Nu_\varepsilon - q_\varepsilon)(\tilde{u} - u_\varepsilon) dx dt \geq 0.$$

This gives

$$\int_Q p_\varepsilon(\tilde{u} - \rho\kappa - \varepsilon q_\varepsilon - u_\varepsilon) dx dt + \int_Q Nu_\varepsilon(\tilde{u} - u_\varepsilon) dx dt - \varepsilon \int_Q q_\varepsilon^2 dx dt \geq \rho \int_Q q_\varepsilon \kappa dx dt,$$

which implies that  $\{q_\varepsilon\}$  is bounded in  $L^{p'}(Q)$ .  $\square$

So  $q_\varepsilon \rightarrow q^*$  weakly in  $L^{p'}(Q)$ , and we can pass to the limit in the necessary conditions since the weak convergence in  $L^{p'}(Q)$  will be coupled with the strong convergence in  $L^p(Q)$ . Thus we have

$$(2.14) \quad \forall y \in K \quad \int_Q (p^* + q^*) \left( \frac{\partial}{\partial t}(y - y^*) - \Delta(y - y^*) \right) dx dt \geq 0.$$

$$(2.15) \quad \forall u \in U_{ad} \quad \int_Q (Nu^* - q^*)(u - u^*) dx dt \geq 0.$$

**Remarks.** (1) These conditions are also sufficient for optimality.

(2) We also notice the completely decoupled character of (2.14), (2.15) and the fact that the Lagrange multiplier  $q^*$  is in  $L^{p'}(Q)$ .

### Example 2

We consider the more difficult case where the state constraint set has a void interior even in  $L^\infty(Q)$ . In the general problem  $(\mathcal{P})$  we may take

$$(2.16) \quad \varphi(x, t) = \psi(x, t) = 0 \quad \text{on } \Sigma,$$

which implies that  $K$  has a void interior in  $C(Q)$ .

We impose a very weak assumption on  $K$  and  $U_{ad}$ , which does not require any interiority conditions:

$$\begin{aligned} & \exists f, g \in L^2(Q), \exists \rho > 0, \exists \tilde{u} \in U_{ad} \text{ such that} \\ (2.17i) \quad & f(x, t) + \rho \leq \tilde{u}(x, t) \leq g(x, t) - \rho \text{ a.e. in } Q, \\ (2.17ii) \quad & \varphi(x, t) \leq y_f(x, t) \leq y_g(x, t) \leq \psi(x, t) \text{ in } \tilde{Q}, \end{aligned}$$

where  $y_f = T(f)$  and  $y_g = T(g)$ . We notice that the pair  $[\tilde{y}, \tilde{u}] = [T(\tilde{u}), \tilde{u}]$  is admissible for  $(\mathcal{P})$ , by comparison.

We define  $y_\kappa = T(\tilde{u} - \rho\kappa)$ ,  $\forall \kappa \in L^\infty(Q)$ ,  $|\kappa|_{L^\infty(Q)} = 1$ . Since it holds

$$f(x, t) \leq \tilde{u}(x, t) - \rho\kappa(x, t) \leq g(x, t) \text{ a.e. in } Q$$

by (2.17i), we get again by comparison

$$y_f(x, t) \leq y_\kappa(x, t) \leq y_g(x, t) \text{ in } \tilde{Q},$$

that is  $y_\kappa \in K$  by (2.17ii). As a consequence, the pair  $[y_\kappa, \tilde{u}]$  is admissible for the problem  $(\mathcal{P}_\varepsilon)$ .

**Proposition 2.4.** *The sequence  $\{q_\varepsilon\}$  is bounded in  $L^1(Q)$  and  $q_\varepsilon \rightarrow q^*$  weakly in  $L^\infty(Q)^* = \mathcal{M}(Q)$  on a generalized subsequence.*

The proof is identical with the proof of Proposition 2.3 by using the above admissible pairs. However, due to the weak convergence and regularity properties, the passage to the limit in the approximate optimality conditions is no more a direct one.

We observe that  $\partial y_\varepsilon / \partial t - \Delta y_\varepsilon = \varepsilon q_\varepsilon + u_\varepsilon + f$ . We use this in (2.11), and then we add (2.11), (2.12) to obtain

$$\int_Q (p_\varepsilon + q_\varepsilon) \left( \frac{\partial y}{\partial t} - \Delta y - \varepsilon q_\varepsilon - u_\varepsilon - f \right) dx dt + \int_Q (Nu_\varepsilon - q_\varepsilon)(u - u_\varepsilon) dx dt \geq 0.$$

Then we have

$$\begin{aligned} (2.18) \quad & \int_Q (p_\varepsilon + q_\varepsilon) \left( \frac{\partial y}{\partial t} - \Delta y - f \right) dx dt - \int_Q p_\varepsilon (\varepsilon q_\varepsilon + u_\varepsilon) dx dt \\ & + \int_Q Nu_\varepsilon (u - u_\varepsilon) dx dt - \int_Q q_\varepsilon u \geq 0 \quad \forall u \in U_{ad}, \forall y \in K. \end{aligned}$$

**Proposition 2.5.** (i) *For any  $u \in U_{ad} \cap L^\infty(Q)$ ,  $y \in K$  such that  $\partial y / \partial t - \Delta y - f \in L^\infty(Q)$ , we have the first order necessary condition for  $(\mathcal{P})$ :*

$$\begin{aligned} (2.19) \quad & \int_Q (p^* + q^*) \left( \frac{\partial y}{\partial t} - \Delta y - f \right) dx dt - \int_Q p^* u^* dx dt \\ & + \int_Q Nu^* (u - u^*) dx dt - \int_Q q^* u \geq 0. \end{aligned}$$

(ii) If the set of admissible regular pairs defined in (i) is dense in the set of all admissible pairs, then (2.19) is also a sufficient optimality condition.

**Proof.** If the test functions  $y$  and  $u$  are regular, then (2.19) is a consequence of (2.18) and of Proposition 2.4.

Conversely, let  $[y, u]$  be any admissible pair for  $(\mathcal{P})$  and  $[y_n, u_n]$  a regular admissible sequence for  $(\mathcal{P})$  such that  $y_n \rightarrow y$  in  $H^{2,1}(Q)$ ,  $u_n \rightarrow u$  in  $L^2(Q)$ . Then, by (2.19) and the fact that  $u_n = \partial y_n / \partial t - \Delta y_n - f$ , we have

$$\int_Q p^* \left( \frac{\partial y_n}{\partial t} - \Delta y_n - f \right) dx dt - \int_Q p^* u^* dx dt + \int_Q N u^* (u_n - u^*) dx dt \geq 0.$$

Passing to the limit  $n \rightarrow \infty$ , we get

$$\int_Q p^* \left( \frac{\partial y}{\partial t} - \Delta y - f \right) dx dt - \int_Q p^* u^* dx dt + \int_Q N u^* (u - u^*) dx dt \geq 0$$

for any admissible  $[y, u]$ . Then we have

$$\int_Q p^* \left( \frac{\partial}{\partial t} (y - y^*) - \Delta (y - y^*) \right) dx dt + \int_Q N u^* (u - u^*) dx dt \geq 0,$$

and integrating by parts, we obtain that the pair  $[y^*, u^*]$  is optimal.  $\square$

**Remarks.** (1) The optimality conditions (2.19) are in a semidecoupled form since only the constraints are decoupled. If  $u^*$  (and consequently  $y^*$ ) is regular, or  $q^*$  is in  $L^2(Q)$ , then the completely decoupled optimality system may be obtained.

(2) The above result cannot be inferred by the classical approach since the Slater assumption is not valid here.

### Example 3

We give an alternative approach to Example 2. The same may be done in Example 1 as well.

Assume that  $X$  is any subspace of the control space  $U$ , possibly not dense. Suppose that there exist  $\tilde{u} \in U$  (not necessarily in  $U_{ad}$ ) and  $\rho > 0$  such that  $\tilde{y} = \mathcal{T}(\tilde{u}) \in K$  and  $\tilde{u} + \rho v \in U_{ad}$  for any  $v \in X$ ,  $|v|_X = 1$ .

The pairs  $[\tilde{y}, \tilde{u} + \rho v]$  are admissible for  $(\mathcal{P}_\varepsilon)$  and, by (2.11), (2.12), we infer

$$\int_Q (p_\varepsilon + q_\varepsilon)(\tilde{u} - \varepsilon q_\varepsilon - u_\varepsilon) dx dt + \int_Q (N u_\varepsilon - q_\varepsilon)(\tilde{u} + \rho v - u_\varepsilon) dx dt \geq 0.$$

Then

$$-\rho \int_Q q_\varepsilon v dx dt \geq - \int_Q N u_\varepsilon (\tilde{u} + \rho v - u_\varepsilon) dx dt + \int_Q p_\varepsilon (\tilde{u} - \varepsilon q_\varepsilon - u_\varepsilon) dx dt$$

for any  $v \in X$ ,  $|v|_X = 1$ , that is  $\{q_\varepsilon\}$  is bounded in  $X^*$ .

If  $X$  equals  $L^p(Q)$ , then we get that  $\{q_\varepsilon\}$  is bounded in  $L^{p'}(Q)$  even in the case where  $K$  has void interior in  $L^\infty(Q)$ . The argument may proceed as in the previous examples.



### 3. An augmented Lagrangian algorithm

As an application of the optimality conditions we present an augmented Lagrangian algorithm for state and control constrained control problems. We make the basic assumption

$$(3.1) \quad q^* \in L^2(Q),$$

which is satisfied in a variety of instances as discussed in Examples 1 and 3. Then the regularity conditions for  $y$  and  $u$  are no longer necessary and the optimality conditions are valid in the decoupled form (2.14), (2.15).

Let  $J(y, u)$  denote the cost function in  $(\mathcal{P})$ . We consider the problem in the abstract form

$$(3.2) \quad \text{Min} \{ J(y, u) \mid y \in K, u \in U_{ad}, y = \mathcal{T}(u) \},$$

where  $\mathcal{T}$  is defined by (1.1)–(1.3). Let  $[y^*, u^*]$  be the unique optimal pair for  $(\mathcal{P})$  (which exists under the usual admissibility hypothesis). There is  $q^* \in L^2(Q)$  such that the optimality conditions (2.14), (2.15) hold, and  $p^*$  is, as before, the solution of (2.4)–(2.6) with  $y_\varepsilon$  replaced by  $y^*$ . Let us define the augmented Lagrangian

$$(3.3) \quad \mathcal{L}_r(y, u, q) = \frac{1}{2} \|y - z_d\|_{L^2(Q)}^2 + \frac{N}{2} \|u\|_{L^2(Q)}^2 \\ + \int_Q q \left( \frac{\partial y}{\partial t} - \Delta y - f - u \right) dx dt + \frac{r}{2} \left\| \frac{\partial y}{\partial t} - \Delta y - f - u \right\|_{L^2(Q)}^2,$$

where  $r$  is a positive penalty parameter. Our algorithm is then as follows:

#### Algorithm.

1. Initialize  $q_0 \in L^2(Q)$  and  $u_{-1} \in L^2(Q)$ .
2. Iterates  $q_n$  and  $u_{n-1}$  being given,

2a. find  $y_n \in K$ , solution of

$$\text{Min}_{y \in K} \left\{ \frac{1}{2} \int_Q (y - z_d)^2 dx dt + \int_Q q_n \left( \frac{\partial y}{\partial t} - \Delta y - f \right) dx dt \right. \\ \left. + \frac{r}{2} \int_Q \left( \frac{\partial y}{\partial t} - \Delta y - f - u_{n-1} \right)^2 dx dt \right\};$$

2b. find  $u_n \in U_{ad}$ , solution of

$$\text{Min}_{u \in U_{ad}} \left\{ \frac{N}{2} \int_Q u^2 dx dt - \int_Q q_n u dx dt \right. \\ \left. + \frac{r}{2} \int_Q \left( \frac{\partial y_n}{\partial t} - \Delta y_n - f - u \right)^2 dx dt \right\}.$$

3. If  $\partial y_n / \partial t - \Delta y_n - f - u_n = 0$ , then STOP, else

4. assign  $q_{n+1} = q_n + \rho(\partial y_n / \partial t - \Delta y_n - f - u_n)$ , where  $0 < \rho_0 \leq \rho \leq r$ , and go to step 2.

**Theorem 3.1.** *Under the above assumptions, we have*

$$(3.4) \quad u_n \rightarrow u^* \quad \text{strongly in } L^2(Q),$$

$$(3.5) \quad y_n \rightarrow y^* \quad \text{strongly in } H^{2,1}(Q),$$

$$(3.6) \quad q_{n+1} - q_n \rightarrow 0 \quad \text{strongly in } L^2(Q).$$

**Proof.** Integrating by parts in (2.14) and noticing that  $\partial y^* / \partial t - \Delta y^* - f - u^* = 0$ , we obtain

$$(3.7) \quad \int_Q (y^* - z_d)(y - y^*) dx dt + \int_Q q^* \left( \frac{\partial}{\partial t}(y - y^*) - \Delta(y - y^*) \right) dx dt \\ + r \int_Q \left( \frac{\partial y^*}{\partial t} - \Delta y^* - f - u^* \right) \left( \frac{\partial}{\partial t}(y - y^*) - \Delta(y - y^*) \right) dx dt \geq 0$$

for every  $y \in K$ . Similarly, from (2.15) we get

$$(3.8) \quad N \int_Q u^*(u - u^*) dx dt - \int_Q q^*(u - u^*) dx dt \\ - r \int_Q \left( \frac{\partial y^*}{\partial t} - \Delta y^* - f - u^* \right) (u - u^*) dx dt \geq 0$$

for every  $u \in U_{ad}$ . Steps 2a and 2b of the algorithm give

$$(3.9) \quad \int_Q (y_n - z_d)(y - y_n) dx dt + \int_Q q_n \left( \frac{\partial}{\partial t}(y - y_n) - \Delta(y - y_n) \right) dx dt \\ + r \int_Q \left( \frac{\partial y_n}{\partial t} - \Delta y_n - f - u_{n+1} \right) \left( \frac{\partial}{\partial t}(y - y_n) - \Delta(y - y_n) \right) dx dt \geq 0$$

and

$$(3.10) \quad N \int_Q u_n(u - u_n) dx dt - \int_Q q_n(u - u_n) dx dt \\ - r \int_Q \left( \frac{\partial y_n}{\partial t} - \Delta y_n - f - u_n \right) (u - u_n) dx dt \geq 0$$

for every  $y \in K$  and  $u \in U_{ad}$ , respectively. Let us substitute  $y = y_n$  to (3.7),  $u = u_n$  to (3.8),  $y = y^*$  to (3.9) and  $u = u^*$  to (3.10), and then add the obtained

inequalities. In this way we get

$$\begin{aligned}
 & \|y_n - y^*\|_{L^2(Q)}^2 + N \|u_n - u^*\|_{L^2(Q)}^2 \\
 & + \int_Q (q_n - q^*) \left( \frac{\partial}{\partial t} (y_n - y^*) - \Delta(y_n - y^*) - (u_n - u^*) \right) dx dt \\
 (3.11) \quad & + r \left\| \frac{\partial}{\partial t} (y_n - y^*) - \Delta(y_n - y^*) - (u_n - u^*) \right\|_{L^2(Q)}^2 \\
 & + r \int_Q \left( \frac{\partial}{\partial t} (y_n - y^*) - \Delta(y_n - y^*) \right) (u_n - u_{n-1}) dx dt \leq 0.
 \end{aligned}$$

Let us denote for a while  $\tilde{y}_n = y_n - y^*$  and  $\tilde{u}_n = u_n - u^*$ . Then we have

$$\begin{aligned}
 & \int_Q \left( \frac{\partial}{\partial t} (y_n - y^*) - \Delta(y_n - y^*) \right) (u_n - u_{n-1}) dx dt \\
 (3.12) \quad & = \int_Q \left( \frac{\partial}{\partial t} (\tilde{y}_n - \tilde{y}_{n-1}) - \Delta(\tilde{y}_n - \tilde{y}_{n-1}) \right) (\tilde{u}_n - \tilde{u}_{n-1}) dx dt \\
 & + \int_Q \left( \frac{\partial \tilde{y}_{n-1}}{\partial t} - \Delta \tilde{y}_{n-1} - \tilde{u}_{n-1} \right) (\tilde{u}_n - \tilde{u}_{n-1}) dx dt \\
 & + \frac{1}{2} \left( \|\tilde{u}_n\|_{L^2(Q)}^2 - \|\tilde{u}_{n-1}\|_{L^2(Q)}^2 - \|\tilde{u}_n - \tilde{u}_{n-1}\|_{L^2(Q)}^2 \right).
 \end{aligned}$$

Let us write (3.10) first with  $n-1$  and  $u = u_n$ , then with  $n$  and  $u = u_{n-1}$ , and add the two inequalities. Then, using the step 4 of the algorithm with  $n-1$ , we get

$$\begin{aligned}
 & r \int_Q \left( \frac{\partial}{\partial t} (\tilde{y}_n - \tilde{y}_{n-1}) - \Delta(\tilde{y}_n - \tilde{y}_{n-1}) \right) (\tilde{u}_n - \tilde{u}_{n-1}) dx dt \\
 (3.13) \quad & \geq (N + r) \|\tilde{u}_n - \tilde{u}_{n-1}\|_{L^2(Q)}^2 \\
 & - \rho \int_Q \left( \frac{\partial y_{n-1}}{\partial t} - \Delta y_{n-1} - f - u_{n-1} \right) (\tilde{u}_n - \tilde{u}_{n-1}) dx dt.
 \end{aligned}$$

Combining (3.12) and (3.13), and noticing again that  $\partial y^* / \partial t - \Delta y^* - f - u^* = 0$ , we obtain

$$\begin{aligned}
 & r \int_Q \left( \frac{\partial}{\partial t} (y_n - y^*) - \Delta(y_n - y^*) \right) (u_n - u_{n-1}) dx dt \\
 & \geq (N + \frac{r}{2}) \|u_n - u_{n-1}\|_{L^2(Q)}^2 + \frac{r}{2} \left( \|u_n - u^*\|_{L^2(Q)}^2 - \|u_{n-1} - u^*\|_{L^2(Q)}^2 \right) \\
 & + (r - \rho) \int_Q \left( \frac{\partial y_{n-1}}{\partial t} - \Delta y_{n-1} - f - u_{n-1} \right) (u_n - u_{n-1}) dx dt.
 \end{aligned}$$

Since  $r - \rho \geq 0$ , this is equivalent to

$$\begin{aligned} & r \int_Q \left( \frac{\partial}{\partial t} (y_n - y^*) - \Delta (y_n - y^*) \right) (u_n - u_{n-1}) \, dx \, dt \\ & \geq (N + \frac{\rho}{2}) \|u_n - u_{n-1}\|_{L^2(Q)}^2 + \frac{r}{2} \left( \|u_n - u^*\|_{L^2(Q)}^2 - \|u_{n-1} - u^*\|_{L^2(Q)}^2 \right) \\ & \quad + \frac{\rho - r}{2} \left\| \frac{\partial y_{n-1}}{\partial t} - \Delta y_{n-1} - f - u_{n-1} \right\|_{L^2(Q)}^2, \end{aligned}$$

and hence, (3.11) becomes

$$\begin{aligned} & \|y_n - y^*\|_{L^2(Q)}^2 + N \|u_n - u^*\|_{L^2(Q)}^2 \\ & + \int_Q (q_n - q^*) \left( \frac{\partial y_n}{\partial t} - \Delta y_n - f - u_n \right) \, dx \, dt + r \left\| \frac{\partial y_n}{\partial t} - \Delta y_n - f - u_n \right\|_{L^2(Q)}^2 \\ & + (N + \frac{\rho}{2}) \|u_n - u_{n-1}\|_{L^2(Q)}^2 + \frac{r}{2} \left( \|u_n - u^*\|_{L^2(Q)}^2 - \|u_{n-1} - u^*\|_{L^2(Q)}^2 \right) \\ & + \frac{\rho - r}{2} \left\| \frac{\partial y_{n-1}}{\partial t} - \Delta y_{n-1} - f - u_{n-1} \right\|_{L^2(Q)}^2 \leq 0. \end{aligned}$$

Let us denote

$$\lambda_n = \frac{r}{2} \|u_{n-1} - u^*\|_{L^2(Q)}^2 + \frac{r - \rho}{2} \left\| \frac{\partial y_{n-1}}{\partial t} - \Delta y_{n-1} - f - u_{n-1} \right\|_{L^2(Q)}^2,$$

so that we have

$$\begin{aligned} & \|y_n - y^*\|_{L^2(Q)}^2 + N \|u_n - u^*\|_{L^2(Q)}^2 \\ (3.14) \quad & + \frac{r + \rho}{2} \left\| \frac{\partial y_n}{\partial t} - \Delta y_n - f - u_n \right\|_{L^2(Q)}^2 + (N + \frac{\rho}{2}) \|u_n - u_{n-1}\|_{L^2(Q)}^2 \\ & \leq \lambda_n - \lambda_{n+1} - \int_Q (q_n - q^*) \left( \frac{\partial y_n}{\partial t} - \Delta y_n - f - u_n \right) \, dx \, dt. \end{aligned}$$

Moreover, let us denote

$$\mu_n = \lambda_n + \frac{1}{2\rho} \|q_n\|_{L^2(Q)}^2 - \frac{1}{\rho} \int_Q q^* q_n \, dx \, dt$$

and use once again the step 4 of the algorithm. Then (3.14) becomes

$$\begin{aligned} & \|y_n - y^*\|_{L^2(Q)}^2 + N \|u_n - u^*\|_{L^2(Q)}^2 \\ & + \frac{r + \rho}{2} \left\| \frac{\partial y_n}{\partial t} - \Delta y_n - f - u_n \right\|_{L^2(Q)}^2 + (N + \frac{\rho}{2}) \|u_n - u_{n-1}\|_{L^2(Q)}^2 \\ & \leq \mu_n - \mu_{n+1} + \frac{\rho}{2} \left\| \frac{\partial y_n}{\partial t} - \Delta y_n - f - u_n \right\|_{L^2(Q)}^2. \end{aligned}$$

This implies

$$(3.15) \quad \|y_n - y^*\|_{L^2(Q)}^2 + N \|u_n - u^*\|_{L^2(Q)}^2 + \frac{r}{2} \left\| \frac{\partial y_n}{\partial t} - \Delta y_n - f - u_n \right\|_{L^2(Q)}^2 \leq \mu_n - \mu_{n+1},$$

so that  $\{\mu_n\}$  is a decreasing sequence. Moreover,  $\lambda_n \geq 0$  and hence

$$\mu_n \geq \frac{1}{2\rho} \left( \|q_n\|_{L^2(Q)}^2 - 2 \int_Q q^* q_n \, dx \, dt \right) \geq -\frac{1}{2\rho} \|q^*\|_{L^2(Q)}^2,$$

that is,  $\{\mu_n\}$  is bounded from below. Therefore, the sequence  $\{\mu_n\}$  is convergent and  $\mu_n - \mu_{n+1} \rightarrow 0$ . The inequality (3.15) and the step 4 of the algorithm imply then the desired convergence results.  $\square$

#### 4. Numerical experiments

We shall now present two example problems and give the numerical results obtained with the augmented Lagrangian algorithm. The discretization of the algorithm is performed by using the finite element method (FEM) with respect to space and the finite difference method (implicit Euler) with respect to time. We consider the problem  $(\mathcal{P})$  with the state system (1.1)–(1.3) and the constraints

$$\begin{aligned} u &\in U_{ad} = \{ u \in L^2(Q) \mid \alpha(x, t) \leq u(x, t) \leq \beta(x, t) \text{ a.e. in } Q \}, \\ y &\in K = \{ y \in H^{2,1}(Q) \mid \varphi(x, t) \leq y(x, t) \leq \psi(x, t) \text{ a.e. in } Q, \\ &\quad y(x, 0) = y_0(x) \text{ a.e. in } \Omega \}. \end{aligned}$$

##### Example 1

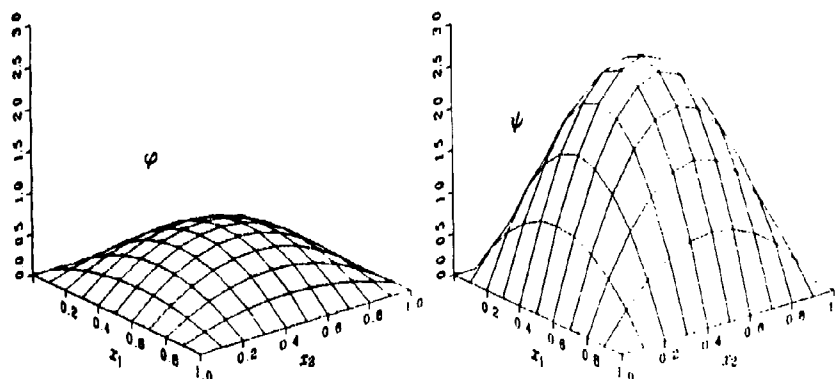
In our first example we have  $\Omega = ]0, 1[ \times ]0, 1[ \subset \mathbb{R}^2$ ,  $T = 1$ ,  $N = 1$  and

$$\begin{aligned} z_d(x, t) &= \sin(\pi x_1) \sin(\pi x_2) e^t, \\ f(x, t) &= (2\pi + 1) \sin(\pi x_1) \sin(\pi x_2) e^t - \sin(3\pi x_1) \sin(3\pi x_2) \left(1 - \frac{t}{2}\right), \\ y_0(x) &= \sin(\pi x_1) \sin(\pi x_2). \end{aligned}$$

Moreover, the constraint functions are

$$\begin{aligned} \alpha &\equiv -1, \quad \beta \equiv 1, \\ \varphi(x, t) &= 4(x_1 - x_1^2)(x_2 - x_2^2)e^t, \\ \psi(x, t) &= 16(x_1 - x_1^2)(x_2 - x_2^2)e^t. \end{aligned}$$

The “desired” state  $z_d$  belongs to  $K$ . Furthermore,  $z_d$  is very close to the upper bound  $\psi$ , and they coincide in the point  $(x_1, x_2) = (0.5, 0.5)$ . Functions  $\varphi$  and  $\psi$  at the time  $t = 1$  are presented in Figure 4.1

FIGURE 4.1. Constraint functions  $\varphi$  and  $\psi$  at the time  $t = 1$ .

The data is chosen such that there is an admissible pair  $[\hat{y}, \hat{u}]$ , where  $\hat{y} = z_u$  and  $\hat{u}(x, t) = \sin(3\pi x_1) \sin(3\pi x_2)(1 - t/2)$ , which, however, is not the optimal pair.

In the space discretization we have  $N_0 = 9 \times 9 = 81$  nodes ( $h = 0.1$ ), and in the time discretization we have  $K_1 = 10$  time levels ( $\Delta t = 0.1$ ). Thus, the dimension of the unknown functions is 810.

As the initial guesses we use  $q_0 \equiv 0$  and  $u_{-1} \equiv 0$ . The minimizer used in the step 2a of the algorithm needs also an initial guess for  $y$ , for which we have chosen  $y(x, t) = y_0(x) \forall t \in ]0, 1]$ . The values of the cost function and the penalty term in this initial point are given in the upper part of Table 4.1

We have tested the algorithm using different values for the penalty parameter  $r$ . The parameter  $\rho$  in step 4 is chosen to be equal to  $r$ . The results are presented in the lower part of Table 4.1. In all four cases, the algorithm found a good solution already in one iteration. In the case  $r = 1$ , the minimizer used in step 2a failed to find a lower value for the cost function after two iterations, and in the other cases,

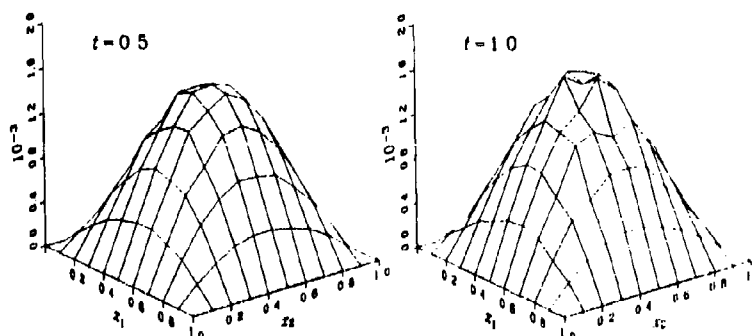
FIGURE 4.2. Control  $u$  at two time levels; Example 1.

TABLE 4.1. Results of the Example 1

		$J(y, u)$	$ \frac{\partial y}{\partial t} - \Delta y - f - u _{L^2(Q)}$
initial values		0.11049	9.9677
$r$	No. of iter.	$J(y, u)$	$ \frac{\partial y}{\partial t} - \Delta y - f - u _{L^2(Q)}$
1	1	$0.18275 \cdot 10^{-3}$	$0.44888 \cdot 10^{-3}$
	2	$0.18237 \cdot 10^{-3}$	$0.77093 \cdot 10^{-3}$
10	1	$0.18337 \cdot 10^{-3}$	$0.90683 \cdot 10^{-5}$
100	1	$0.18344 \cdot 10^{-3}$	$0.28708 \cdot 10^{-6}$
1000	1	$0.18343 \cdot 10^{-3}$	$0.57937 \cdot 10^{-7}$

TABLE 4.2. Results of the Example 2

$r = 1$			$r = 10$		
No. of iter.	$J(y, u)$	$ \frac{\partial y}{\partial t} - \Delta y - f - u $	No. of iter.	$J(y, u)$	$ \frac{\partial y}{\partial t} - \Delta y - f - u $
0	0.17204	0.58508	0	0.17204	0.58508
1	0.21115	0.27934	1	0.27574	0.11628
2	0.21037	0.27651	2	0.27666	0.11291
3	0.20906	0.27148	3	0.27875	0.10585
4	0.20854	0.26942	4	0.27687	0.11053
5	0.20808	0.26751	5	0.27873	0.10416
6	0.20844	0.26910	6	0.27755	0.10954
7	0.20838	0.26876	7	0.27932	0.10377
8	0.20802	0.26724	8	0.27953	0.10175
9	0.20796	0.26693	9	0.27851	0.10538
10	0.20828	0.26833	10	0.27893	0.10331

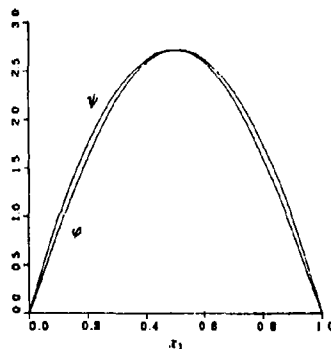
$r = 100$			$r = 1000$		
No. of iter.	$J(y, u)$	$ \frac{\partial y}{\partial t} - \Delta y - f - u $	No. of iter.	$J(y, u)$	$ \frac{\partial y}{\partial t} - \Delta y - f - u $
0	0.17204	0.58508	0	0.17204	0.58508
1	0.28451	0.10896	1	0.28540	0.10855
2	0.28618	0.10491	2	0.28703	0.10453
3	0.29196	0.09399	3	0.28661	0.10539
4	0.28772	0.10127	4	0.28687	0.10575
5	0.29129	0.09479	5	0.28670	0.10754
6	0.28837	0.10068	6	0.28721	0.10687
7	0.28726	0.10438	7	0.28764	0.10709
8	0.29046	0.09790	8	0.28795	0.10760
9	0.29063	0.09829	9	0.28861	0.10730
10	0.29101	0.09798	10	0.28990	0.10645

this happened already after one iteration.

As could be expected, the larger the parameter  $r$  is, the better the constraint  $\partial y / \partial t - \Delta y = f + u$  is satisfied. On the other hand, with smaller  $r$  the value of the cost function seems to get somewhat lower. In Figure 4.2 is presented the control  $u$  obtained in the case  $r = 1$ .

### Example 2

FIGURE 4.3. Constraint functions  $\varphi$  and  $\psi$  when  $x_2 = \frac{1}{2}$  and  $t = 1$ .



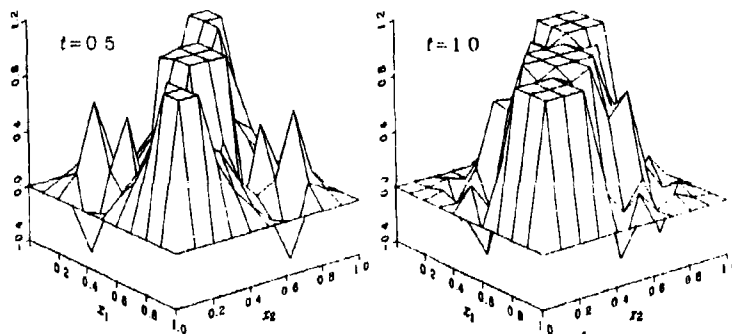
Our second example is a variant of the first one. The problem data is the same, except the following modifications have been made:

Now we have  $z_d \equiv 1/2$ , and

$$\varphi(x, t) = \sin(\pi x_1) \sin(\pi x_2) e^t,$$

$$\psi(x, t) = 16(x_1 - x_1^2)(x_2 - x_2^2) e^t.$$

FIGURE 4.4. Control  $u$  at two time levels: Example 2.





In this case,  $z_d$  does not belong to  $K$ . The constraint functions  $\varphi$  and  $\psi$  are chosen such that the set  $K$  is very "small": see Figure 4.3. Moreover, there is no interior point at  $(x_1, x_2) = (0.5, 0.5)$ .

The pair  $[\hat{y}, \hat{u}]$ , where  $\hat{y} = \varphi$  and  $\hat{u}(x, t) = \sin(3\pi x_1) \sin(3\pi x_2)(1 - t/2)$  is admissible for this problem. The discretization is the same as before, as well as the initial guesses for  $q$  and  $u$ . The initial guess for  $y$ , however, is different, since  $y(x, t) = y_0(x) \forall t \in [0, 1]$  does not belong to  $K$ ; in this case we have used  $y \equiv \varphi$ .

Again, we have tested the algorithm using different values for  $r$ . The step parameter  $\rho$  is chosen to be equal to  $r$ . The results are presented in Table 4.2 (With the iteration '0' we mean the values at the initial point.)

In this example, the algorithm did not stop in the early iterations, as was the case in the previous example. However, due to the large computation time, only ten iterations have been calculated. In all four cases, the convergence is quite slow after the first iteration. If we look at the constraint  $\partial y / \partial t - \Delta y = f + u$ , the parameter  $r = 100$  seems to give the "best" convergence. (The algorithm was tried also for  $r = 10^4$  and  $r = 10^5$ ; the results were similar to the case  $r = 1000$ .) In Figure 4.4 is presented the control  $u$  obtained in the case  $r = 100$ .

## References

1. M. Bergounioux, *Problèmes de contrôle avec contraintes sur l'état*, C. R. Acad. Sci. Paris, Sér. I Math. **310** (1990), 391–396.
2. M. Bergounioux, "Optimal control of parabolic problems with state constraints: a penalization method for optimality conditions", Rapport de Recherche, Université d'Orléans, 1991.
3. M. Bergounioux, *A penalization method for optimal control of elliptic problems with state constraints*, SIAM J. Control and Optimization **30** (1992), 305–323.
4. M. Bergounioux, T. Männikkö, D. Tiba, "On non-qualified parabolic control problems", Preprint 148, University of Jyväskylä, Department of Mathematics, 1992.
5. J.F. Bonnans, E. Casas, *On the choice of the function spaces for some state constrained control problems*, Numer. Funct. Anal. and Optim. **7** (1984-85), 333–348.
6. J.F. Bonnans, E. Casas, *Optimal control of semilinear multistate systems with state constraints* SIAM J. Control and Optimization **27** (1989), 446–455.

M. Bergounioux  
Département de Mathématiques et d'Informatique  
U.F.R. Sciences Université d'Orléans  
B.P. 6779  
F-45067 Orléans Cedex 2, France

T. Männikkö  
Department of Mathematics  
University of Jyväskylä  
P.O. Box 35  
FIN-40351 Jyväskylä, Finland

D. Tiba  
Institute of Mathematics of the Romanian Academy  
P.O. Box 1-764  
RO-70700 Bucharest, Romania

## CONVERGENCE OF TRAJECTORIES FOR A CONTROLLED VISCOUS BURGERS' EQUATION

CHRISTOPHER I. BYRNES, DAVID S. GILLIAM, AND VICTOR I. SHUBOV

Department of Systems Science and Mathematics  
Washington University

Department of Mathematics  
Texas Tech University

**ABSTRACT.** In this paper we consider a boundary control problem for Burgers' equation on a finite interval. The controls enter as gain parameters in the boundary conditions as in [1, 2]. The uncontrolled problem is obtained by equating the control parameters to zero while the zero dynamics system is obtained by equating the control parameters to infinity, or (intuitively) as the "high gain limit" of the system as the gains approach plus infinity. The main result of the paper is a nonlinear enhancement of the classical root locus result which states that the trajectories of the closed loop system converge to the trajectories of the zero dynamics system as the gains are increased to infinity.

*1991 Mathematics Subject Classification.* 93B05, 93B28, 93B52

*Key words and phrases.* Boundary control, nonlinear distributed parameter systems, zero dynamics, convergence of trajectories.

### 1. Introduction

The need to control spatially distributed systems or systems in which rigid dynamics are coupled with the effects of fluid flow and flexible structural modes has sparked a tremendous effort aimed at the stabilization or control of distributed parameter systems. For example, the existence and analysis of boundary feedback control laws which stabilize classes of linear, distributed parameter systems has attracted the interest of many researchers, using a variety of techniques drawn from both state-space and frequency domain methods. While the differences between these two methodologies are certainly heightened in the distributed parameter case, one of the elegant aspects of linear geometric control theory is its capacity for giving intuitive state-space formulations of many frequency domain concepts and constructs. It is also worth noting that the extension of geometric methods to

---

This work was supported in part by grants from NSF, AFOSR and Texas Advanced Research Program.

lumped nonlinear systems at the beginning of the last decade spearheaded a virtual renaissance in nonlinear control theory, yielding systematic methodologies for the design of feedback laws while retaining important intuitive aspects of classical state-space and frequency design methods.

One of the important feedback design methods of classical automatic control is root locus theory, based on the observation that in the frequency domain output the closed loop poles of a system vary from the open loop poles to the open loop zeros as the gain is increased from zero to infinity. Successfully exploited for decades for finite dimensional systems, this fundamental method has been extended to the nonlinear finite dimensional case ([4]), where it is known that as certain gain parameters are tuned the closed-loop trajectories approach the trajectories of the zero dynamics. On the other hand, root locus methods have also recently been extended to the infinite dimensional case in [4] where a fairly complete analog of finite dimensional root locus theory is developed for a class of parabolic boundary control problems in which the inputs and outputs occur through certain boundary operators and a closed loop system is obtained by employing a proportional error feedback law,  $u = -ky$ . In this case, in [4] it is shown that the infinitely many closed loop poles vary from the open loop poles to the open loop zeros as the gain is increased from zero to plus or minus infinity, depending on the sign of the instantaneous gain. Defining the zero dynamics to be the system obtained by constraining the output  $y$  to zero, or equivalently, as the system obtained in the high gain limit, it is possible to enhance many of the classical results on stabilization of minimum phase systems (i.e., systems with exponentially stable zero dynamics). In particular, in [4] it is shown that the one parameter family of closed loop spatial operators  $A_k$  — the analog of  $(A + BkC)$  in classical finite dimensional theory — form an analytic family, in the gain parameter  $k$ , of unbounded operators, in Hilbert state space, in the sense of norm resolvent convergence (cf. Kato [11]). This result together with a generalization the Trotter-Kato theorem provides a simple proof that the semigroups  $S_k(t)$ , with infinitesimal generators  $A_k$ , converge in the uniform operator topology to the semigroup generated by the zero dynamics,  $S_\infty(t)$ .

The main result of the present paper is to provide a nonlinear enhancement of the root locus results obtained for linear distributed parameter systems in [4] for a boundary controlled, viscous Burgers' equation. Explicitly, we show that the closed-loop trajectories approach the trajectories of the zero dynamics as certain gain parameters are tuned. Since the zero dynamics is exponentially stable, this design method also provides a class of stabilizing feedback laws. The most important result used in the proofs is that the linear part of the closed loop spatial Burgers' operator form a holomorphic family in the sense of norm resolvent convergence and that the nonlinear term is a Lipschitz operator with respect to the Hilbert scale norm generated by the square root of the linear part. This paper represents a special case of the results found in [6] where we consider the same system but with a nonzero external forcing term  $f(x, t)$ . In this case we do not obtain exactly the same result. Namely we do not obtain exponential stability. Further in [6] we

show that for nonzero forcing terms which are independent of time, there exists, for every nonzero value of the gain parameter, a local attractor. It is also shown that as the gains tend to plus infinity the local attractors converge to the single global asymptotically stable equilibrium for the zero dynamics system. In [7] we consider the problem of convergence of trajectories using a priori estimates and Galerkin approximations for small  $L^2$  initial data. This, of course, represents a marked improvement over the results for small  $H^1$  initial data. We are not able to extend our semigroup methods used for  $H^1$  initial data to this more general case. On the other hand we obtain stronger results in [6] which is to be expected for smoother initial data.

## 2. The Burgers' System and Zero Dynamics

Consider the uncontrolled viscous Burgers' system

$$(2.1) \quad \begin{aligned} w_t(x, t) + \epsilon A_0 w(x, t) &= F(w(x, t)), \quad x \in (0, 1), \quad t > 0 \\ 0 &= -w_x(0, t), \quad 0 = w_x(1, t), \quad w(x, 0) = \phi(x) \end{aligned}$$

where  $A_0$  is the unbounded operator  $A = -\frac{d^2}{dx^2}$  with dense domain in  $L^2(0, 1)$

$$D(A_0) = \{f \in H^2 \mid f'(0) = f'(1) = 0\}$$

( $' = d/dx$ ) and

$$F(w) = -\frac{dw}{dx}w.$$

The Controlled Burgers' system is defined by

$$(2.2) \quad \begin{aligned} w_t(x, t) + \epsilon A_k w(x, t) &= F(w(x, t)), \quad x \in (0, 1), \quad t > 0 \\ -w_x(0, t) + k_0 w(0, t) &= 0, \quad w_x(1, t) + k_1 w(1, t) = 0, \\ w(x, 0) &= \phi(x). \end{aligned}$$

where  $A_k = A$  with domain  $\mathcal{D}(A_k) = \{f \in H^2(0, 1) : f'(0) - k_0 f(0) = 0, f'(1) + k_1 f(1) = 0\}$  and  $k = (k_0, k_1)$ . Note that for  $k_0 = k_1 = 0$ , the uncontrolled system (2.1) is obtained and the parameters  $k_0$  and  $k_1$  with  $k_0 + k_1 > 0$  are considered as boundary controls or gain parameters.

The zero dynamics for the controlled Burgers' system is the system obtained for infinite value of the gain parameters. This corresponds to the system

$$(2.3) \quad \begin{aligned} w_t(x, t) + \epsilon A_\infty w(x, t) &= F(w(x, t)), \quad x \in (0, 1), \quad t > 0 \\ w(1, t) &= 0, \quad w(0, t) = 0, \quad w(x, 0) = \phi(x) \end{aligned}$$

where  $A_\infty = A$  with domain  $\mathcal{D}(A_\infty) = \{f \in H^2(0, 1) : f(0) = 0, f(1) = 0\}$ .

### 3. The Linearization about Zero

The linearization about zero of (2.2) is the controlled heat equation

$$(3.1) \quad \begin{aligned} w_t(x, t) + \epsilon A_k w(x, t), \quad x \in (0, 1), \quad t > 0 \\ -w_x(0, t) + k_0 w(0, t) &= 0 \\ w_x(1, t) + k_1 w(1, t) &= 0 \\ w(x, 0) &= \phi(x). \end{aligned}$$

For  $k_0 = k_1 = 0$ , neither (2.2) nor the linearization (3.1) is asymptotically stable, but for  $k_0 + k_1 > 0$  the linearization (3.1) is asymptotically stable. In fact,  $A_k$  is a strictly positive selfadjoint operator for  $k_0 + k_1 > 0$ , so that by the Lumer-Phillips Theorem  $(-\epsilon A_k)$  generates a contraction semigroup  $S_k(t)$  satisfying

$$\|S_k(t)\| \leq e^{-\epsilon \lambda_1(k)t}$$

where  $\lambda_1(k)$  is the first eigenvalue of  $A_k$ . In this case, it is easy to show that

$$0 > \lambda_1(k) \rightarrow \pi^2, \quad k_0, k_1 \rightarrow \infty.$$

Thus the system (3.1) has solution  $w(t)$  satisfying

$$\|w(t)\| = \|S_k(t)\phi\| \leq e^{-\epsilon \lambda_1(k)t} \|\phi\|$$

where  $\|\cdot\|$  is the norm in  $L^2(0, 1)$  induced by the inner product

$$(\phi, \psi) = \int_0^1 \phi(x) \overline{\psi(x)} dx.$$

The spectrum of  $A_k$  is given by positive numbers  $\lambda = \mu^2$  as the roots of the characteristic equation

$$(k_0 k_1 - \mu^2) \frac{\sin(\mu)}{\mu} + (k_0 + k_1) \cos(\mu) = 0.$$

This equation has infinitely many zeros  $\{\mu_j\}_{j=1}^\infty$  satisfying

$$(j-1)\pi < \mu_j < j\pi, \quad j = 1, 2, \dots.$$

Corresponding to the eigenvalues  $\lambda_j = \mu_j^2$ , we have the complete orthonormal system of eigenfunctions

$$\psi_j(x) = \kappa_j \left( k_0 \frac{\sin(\mu_j x)}{\mu_j} + \cos(\mu_j x) \right), \quad j = 1, 2, \dots$$

in  $L^2(0, 1)$  where

$$\kappa_j = \sqrt{\frac{2\mu_j^2(k_1^2 + \mu_j^2)}{(k_0^2 + \mu_j^2)(k_1^2 + \mu_j^2) + (k_0 + k_1)(\mu_j^2 + k_0 k_1)}}.$$

is the normalization constant.

The operators  $A_k$  define an infinite scale of Hilbert spaces  $\mathcal{H}^\alpha$  ( $\alpha \in \mathcal{R}$ ). If  $\alpha \geq 0$  then  $\mathcal{H}^\alpha$  consists of vectors  $\phi \in \mathcal{H}^0 = L^2(0, 1)$  such that

$$(3.2) \quad \|\phi\|_\alpha = \left( \sum_{j=1}^{\infty} \lambda_j^\alpha (\phi, \psi_j)^2 \right)^{1/2} < \infty.$$

The space  $\mathcal{H}^\alpha$  is the domain of the operator  $A_k^{\alpha/2}$  and the inner product in this space is given by

$$(\phi, \psi)_\alpha = (A_k^{\alpha/2} \phi, A_k^{\alpha/2} \psi)$$

which is the same as (3.2) for  $\psi = \phi$ . The operator  $A_k^{\alpha/2}$  is defined on  $\mathcal{H}^\alpha$  by the formula

$$A_k^{\alpha/2} \phi = \sum_{j=1}^{\infty} \lambda_j^{\alpha/2} (\phi, \psi_j) \psi_j.$$

Using the fact that  $\mathcal{D}(A_k)$  is a core in  $\mathcal{D}(A_k^{1/2})$  it is easy to show that the norm in  $\mathcal{H}^1$  can be written as

$$(3.3) \quad \|\varphi\|_1^2 = \int_0^1 |\varphi_x|^2 dx + k_0 |\varphi(0)|^2 + k_1 |\varphi(1)|^2.$$

The spaces  $\mathcal{H}^\alpha$  have the following properties:

1. If  $\beta > \alpha$  then  $\mathcal{H}^\beta \subset \mathcal{H}^\alpha$  and  $\|\phi\|_\alpha \leq \lambda_1(k)^{(\alpha-\beta)/2} \|\phi\|_\beta$  for all  $\phi \in \mathcal{H}^\beta$ ;  $\mathcal{H}^\beta$  is dense in  $\mathcal{H}^\alpha$ .
2. The imbedding  $\mathcal{H}^\beta \subset \mathcal{H}^\alpha$  is compact.

The semigroup  $S_k(t)$  is given by

$$S_k(t)f = \sum_{j=1}^{\infty} e^{-\lambda_j(k)t} f_j \psi_j$$

from which it is easy to show that  $\|S_k(t)\| \leq e^{-\lambda_1(k)t}$ .

#### 4. Global Existence for Small Initial Data

The proof of the following result concerning global (in time) existence and exponential stability of solutions for small initial data in  $H^1(0, 1)$  follows exactly the proof given in Theorem 3.3.3, Theorem 6.3.3, page 199 of [10] and Lemma 5.6.7, page 159 of [12]. A more general situation is considered in the paper [6] where the controlled Burgers' equation has an additional forcing term  $f(x, t)$ . In this case we do not obtain the exponential convergence of the trajectories to zero as  $t$  goes to infinity. Nevertheless, even in this much more general situation we are able to obtain convergence of the trajectories for small forcing terms. Furthermore, if the forcing term is independent of  $t$ , then we establish the exist in [6] of local attractors in  $H^1$  for each positive value of the gain parameters and prove that, in an

appropriate sense, the local attractors converge to the single, global asymptotically stable equilibrium for the zero dynamics.

**Theorem 4.1.** *For  $k = (k_0, k_1)$ ,  $k_0, k_1 \geq 0$ ,  $k_0 + k_1 > 0$  and any  $\beta$  satisfying  $0 < \beta < \lambda_1(k) < \pi^2$ , there exists  $\rho > 0$  such that for  $\phi \in H^1(0, 1)$  with*

$$\|\phi\|_1 \leq \rho/2$$

*there is a unique solution  $z(t) \in H^1$  for all  $t > 0$  of*

$$z_t(t) + \epsilon A_k z(t) = F(z(t)), \quad z(0) = \phi$$

*and, moreover, the solution satisfies*

$$(4.1) \quad \|z(t)\|_1 \leq 2e^{-c\beta t} \|\phi\|_1, \quad t \geq 0.$$

*With  $c = \sqrt{2}(1 + \lambda_1(k)^{-1})^{1/2}$ ,  $\rho$  can be taken to be any number such that*

$$(4.2) \quad \rho \frac{c}{\epsilon(\lambda_1(k) - \beta)} \left( \lambda_1(k)^{1/2} + \sqrt{\frac{\pi}{2}} (\lambda_1(k) - \beta)^{1/2} \right) < \frac{1}{2}.$$

The following simple estimates are used in establishing the theorem and also in the sequel. We state them without proof and refer to [6].

**Lemma 4.1.** *For each  $\varphi \in \mathcal{H}^1$  we have*

$$\|\varphi\|^2 \leq \lambda_1(k)^{-1} \|\varphi\|_1^2,$$

*where  $0 < \lambda_1(k) < \pi^2$  is the first eigenvalue of  $A_k$ . Furthermore, the norm (3.3) is equivalent to the  $H^1(0, 1)$  Sobolev norm and, therefore  $\mathcal{H}^1 = H^1(0, 1)$ .*

**Lemma 4.2.** *For  $z \in H^1(0, 1)$  we have the estimates*

$$\|z\|_\infty \leq \sqrt{2} \|z\|_{H^1}, \quad \|z\|_\infty \leq c \|z\|_1$$

*where*

$$\|z\|_{H^1}^2 = \|z\|^2 + \|z_x\|^2, \quad \|z\|_\infty = \operatorname{ess\,sup}_{x \in [0, 1]} |z(x)|$$

*and*

$$(4.3) \quad c = \sqrt{2}(1 + \lambda_1(k)^{-1})^{1/2}.$$

With these estimates it is easy to see (cf. [10]) that  $F$  is locally Lipschitz, i.e., for  $\|z_1\|_1, \|z_2\|_1 < M$ , there exists a constant,  $C$  depending only on  $M$ , and  $k$  but otherwise independent of  $z_1, z_2$  such that

$$\|F(z_1) - F(z_2)\| \leq C \|z_1 - z_2\|_1.$$

This allows us to apply the basic local existence result found in [10] and [12].



**Lemma 4.3.** For  $z \in \mathcal{H}^1$  and  $F(z) = -zz_x$  we have

$$(4.4) \quad \|F(z)\| \leq c(\|z\|_1)^2$$

where  $c$  is defined in (4.3).

**Lemma 4.4.** The following estimates hold

1. If  $g \in H^\alpha$  then for all  $t > 0$ , we have

$$\|S_k(t)g\|_\alpha \leq e^{-\epsilon\lambda_1(k)t} \|g\|_\alpha.$$

2. For  $g \in L^2$  and all  $t > 0$

$$(4.5) \quad \|S_k(t)g\|_\alpha \leq \left( \left( \frac{\alpha}{2\epsilon t} \right)^{\alpha/2} + \lambda_1(k)^{\alpha/2} \right) e^{-\epsilon\lambda_1(k)t} \|g\|.$$

## 5. Convergence of Trajectories for Burgers' Equation

The main result of this paper is the following theorem where throughout the remainder of the paper we assume the hypotheses of Theorem 2.1.

**Theorem 5.1.** Let  $z_k$  and  $z_\infty$  denote the solutions of (2.2) and (2.3), respectively for the same initial data  $\phi \in \mathcal{H}^1$  with

$$\|\phi\|_1 \leq \rho/2.$$

Then for arbitrary  $\delta > 0$ , there exists a  $K > 0$  so that for  $k_0, k_1 > K$  we have

$$\|z_k(\cdot, t) - z_\infty(\cdot, t)\|_\infty \leq \delta e^{-\epsilon t}$$

for all  $t \geq 0$ .

In order to establish both convergence in the spatial variable for increasing  $k_0$  and  $k_1$  as well as exponential convergence uniformly in time we introduce a shifted semigroup and infinitesimal generator. For  $0 < \beta < \lambda_1(k)$  let

$$S_{k,\beta}(t) = e^{\epsilon\beta t} S_k(t) = e^{-\epsilon(A_k - \beta I)t}, \quad S_{\infty,\beta}(t) = e^{\epsilon\beta t} S_\infty(t) = e^{-\epsilon(A_\infty - \beta I)t}$$

with corresponding infinitesimal generators given by

$$-\epsilon A_{k,\beta} = \epsilon(-A_k + \beta I), \quad -\epsilon A_{\infty,\beta} = \epsilon(-A_\infty + \beta I).$$

The family of operators  $A_{k,\beta}$  generate an infinite scale of Hilbert spaces which are related to, indeed equal as functions to, the Hilbert scale  $\mathcal{H}^\alpha$  generated by  $A_k$ . Namely, with

$$C_\beta = \frac{\lambda_1(k) - \beta}{\lambda_1(k)}$$

we have

$$C_\beta^{\alpha/2} \|A_k^{\alpha/2} \phi\| \leq \|A_{k,\beta}^{\alpha/2} \phi\| \leq \|A_k^{\alpha/2} \phi\|.$$

The proof is straightforward: First note that

$$\|A_{k,\beta}^{\alpha/2}\phi\|^2 = \|(A_k - \beta I)^{\alpha/2}\phi\|^2 = \sum_{j=1}^{\infty} (\lambda_j - \beta)^{\alpha} |\phi_j|^2 \leq \sum_{j=1}^{\infty} (\lambda_j)^{\alpha} |\phi_j|^2 = \|A_k^{\alpha/2}\phi\|^2$$

and with the definition of  $C_{\beta}$  above we have  $(\lambda_j - \beta) > C_{\beta}\lambda_j$  for all  $j$ , so that

$$\begin{aligned} C_{\beta}^{\alpha} \|A_k^{\alpha/2}\phi\|^2 &= C_{\beta}^{\alpha} \sum_{j=1}^{\infty} (\lambda_j)^{\alpha} |\phi_j|^2 \\ &\leq \sum_{j=1}^{\infty} (\lambda_j - \beta)^{\alpha} |\phi_j|^2 = \|(A_k - \beta I)^{\alpha/2}\phi\|^2 = \|A_{k,\beta}^{\alpha/2}\phi\|^2. \end{aligned}$$

Now recall that a classical solution of (2.2) satisfies the variation of parameters formula for  $0 < k_0 + k_1 < \infty$

$$z_k(t) = S_k(t)\phi + \int_0^t S_k(t-\tau)F(z_k(\tau))d\tau,$$

as does the solution of the zero dynamics (2.3)

$$z_{\infty}(t) = S_{\infty}(t)\phi + \int_0^t S_{\infty}(t-\tau)F(z_{\infty}(\tau))d\tau.$$

Also from Lemma 4.2, we have the pointwise estimate

$$|z_k(x, t) - z_{\infty}(x, t)| \leq c \|z_k(t) - z_{\infty}(t)\|_1$$

with  $c$  defined in (4.3).

So to compute a pointwise estimate for the difference of the solutions we need only consider the  $\mathcal{H}^1$  norm of the difference.

$$\begin{aligned} e^{\epsilon\beta t} \|z_k(\cdot, t) - z_{\infty}(\cdot, t)\|_1 &= e^{\epsilon\beta t} \|A_k^{1/2}(z_k - z_{\infty})\| \leq e^{\epsilon\beta t} \|A_k^{1/2}[S_k(t)\phi - S_{\infty}(t)\phi]\| \\ &\quad + e^{\epsilon\beta t} \int_0^t \|A_k^{1/2}[S_k(t-\tau)F(z_k(\tau)) - S_{\infty}(t-\tau)F(z_{\infty}(\tau))]\| d\tau \\ &= \|A_k^{1/2}[S_{k,\beta}(t)\phi - S_{\infty,\beta}(t)\phi]\| + \int_0^t \|A_k^{1/2}[S_{k,\beta}(t-\tau)e^{\epsilon\beta\tau}F(z_k(\tau)) \\ &\quad - S_{\infty,\beta}(t-\tau)e^{\epsilon\beta\tau}F(z_{\infty}(\tau))]\| d\tau \\ &\leq C_{\beta}^{-1/2} \|A_{k,\beta}^{1/2}[S_{k,\beta}(t)\phi - S_{\infty,\beta}(t)\phi]\| + C_{\beta}^{-1/2} \left\| \int_0^t A_{k,\beta}^{1/2} S_{k,\beta}(t-\tau)e^{\epsilon\beta\tau}F(z_k(\tau)) \right. \\ &\quad \left. - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t-\tau)e^{\epsilon\beta\tau}F(z_{\infty}(\tau))d\tau \right\| \end{aligned}$$

$$+ C_\beta^{-1/2} \left\| \int_0^t [A_{k,\beta}^{1/2} - A_{\infty,\beta}^{1/2}] S_{\infty,\beta}(t-\tau) e^{\epsilon\beta\tau} F(z_\infty(\tau)) d\tau \right\|$$

$$\equiv I + II + III.$$

For the term  $II$  we have the estimates

$$C_\beta^{1/2} II \leq \int_0^t \|A_{k,\beta}^{1/2} S_{k,\beta}(t-\tau) e^{\epsilon\beta\tau} (F(z_k(\tau)) - F(z_\infty(\tau)))\| d\tau$$

$$+ \left\| \int_0^t [A_{k,\beta}^{1/2} S_{k,\beta}(t-\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t-\tau)] e^{\epsilon\beta\tau} F(z_\infty(\tau)) d\tau \right\|$$

$$\equiv IIa + IIb.$$

For the term  $IIa$  we need the estimates

$$e^{\epsilon\beta\tau} \|F(z_k) - F(z_\infty)\| = e^{\epsilon\beta\tau} \|z_k z_{kx} - z_\infty z_{\infty x}\|$$

$$= e^{\epsilon\beta\tau} \|z_{kx}(z_k - z_\infty) + z_\infty(z_{kx} - z_{\infty x})\|$$

$$\leq e^{\epsilon\beta\tau} (\|z_{kx}\| \|z_k - z_\infty\|_\infty + \|z_\infty\|_\infty \|z_{kx} - z_{\infty x}\|)$$

$$\leq e^{\epsilon\beta\tau} (\|z_k\|_1 e^{\epsilon\beta\tau} \|z_k - z_\infty\|_1 + e^{\epsilon\beta\tau} \|z_\infty\|_1) \|z_k - z_\infty\|_1$$

$$\leq e^{\epsilon\beta\tau} (\|z_k\|_1 + \|z_\infty\|_1) \|z_k - z_\infty\|_1$$

$$\leq e^{\epsilon\beta\tau} 2e^{-\epsilon\beta\tau} \|\phi\|_1 \|z_k - z_\infty\|_1 \leq c\rho e^{\epsilon\beta\tau} \|z_k - z_\infty\|_1$$

where we have used (4.1). Now using (4.5) we obtain

$$IIa \leq c\rho \int_0^t \left( \frac{1}{\sqrt{2\epsilon(t-\tau)}} + \lambda_1(k)^{1/2} \right) e^{-\epsilon\lambda_1(k)(t-\tau)} e^{\epsilon\beta\tau} \|z_k - z_\infty\|_1 d\tau$$

$$\leq (1/2) \sup_{0 < \tau < t} e^{\epsilon\beta\tau} \|z_k - z_\infty\|_1$$

where the last inequality follows from our choice of  $\rho$  in (4.2). Now defining

$$\omega(t) = \sup_{0 < \tau < t} e^{\epsilon\beta\tau} \|z_k - z_\infty\|_1.$$

As in the proof of Theorem 4.1 it has been shown that

$$(5.1) \quad e^{\epsilon\beta t} \|z_k(\cdot, t) - z_\infty(\cdot, t)\|_1 \leq I + IIb + III + \frac{1}{2}\omega(t)$$

and we need only show that the first three terms on the right can be made less than  $\delta > 0$  for  $k$  sufficiently large, uniformly in  $t$ . Then we will have

$$\frac{1}{2}\omega(t) \leq \delta.$$

The term  $I$  is considered in detail in Lemmas 5.1 to 5.4. The term  $IIb$  is considered in Lemma 5.6 and the last term  $III$  is examined in Lemma 5.5.

As a consequence of the main theorem of [4] (see also [11], Theorem 1.14 and Example 1.15, page 374) we have that the negative selfadjoint operators  $(-A_k)$  form a holomorphic family in  $k_0, k_1 \in [0, \infty]$  with  $k_0 + k_1 > 0$  in the sense of Kato ([11], Theorem 2.25, page 206 and Theorem 1.3 and Example 1.4, page 367). Therefore, defining

$$R_k(\lambda) = (\lambda I + A_k)^{-1}$$

for any  $\lambda \notin (-\infty, 0)$  we have

$$\|R_k(\lambda) - R_\infty(\lambda)\| \rightarrow 0, \quad k \rightarrow \infty$$

where  $k \rightarrow \infty$  means  $k_0, k_1 \rightarrow \infty$ . In one form or another most of the following results repose on this strong statement concerning the fact that the resolvents converge in the uniform operator topology as  $k_0$  and  $k_1$  tends to infinity.

Let us now recall some facts concerning the resolvents.

**Lemma 5.1.** 1. For every  $f \in L^2$ ,

$$\|sR_k(s)f - f\| \rightarrow 0, \quad s \rightarrow +\infty$$

and

$$\|sR_k(s)\| \leq 1.$$

It also follows that

$$\|s^2R_k(s)^2f - f\| \rightarrow 0, \quad s \rightarrow +\infty.$$

2. For all  $t \geq 0$  and  $s > 0$

$$\|R_k(s)(S_\infty(t) - S_k(t))R_\infty(s)\| \leq te^{-\epsilon\lambda_1(k)t}\|R_k(s) - R_\infty(s)\|.$$

3. For all  $t \geq 0$  and  $s > 0$

$$\begin{aligned} [S_k(t) - S_\infty(t)]R_\infty(s)^2 &= R_k(s)[S_k(t) - S_\infty(t)]R_\infty(s) \\ &\quad + [R_k(s) - R_\infty(s)]S_\infty(t)R_\infty(s) - S_k(t)[R_k(s) - R_\infty(s)]R_\infty(s). \end{aligned}$$

4. For all  $t \geq 0$  and  $s > 0$

$$\|[S_k(t) - S_\infty(t)]R_\infty^2(s)\| \leq (t+2)e^{-\epsilon\lambda_1(k)t}\|R_k(s) - R_\infty(s)\|$$

where we recall that  $\lambda_1(k) < \lambda_1(\infty)$ .

**Proof:** Part 1 is a well known consequence of the fact that  $-A_k$  generates a contraction semigroup (cf. [12]). Part 2 can be found in [11] (page 501, Theorem 2.14). Part 3 is simple algebra obtained by adding and subtracting certain terms. Part 4 follows from parts 1 through 3.  $\square$

Again we recall that the standing hypotheses are those of Theorem 2.1.

**Lemma 5.2.** *For every  $\delta > 0$  we can find a  $K$  for which  $k_0, k_1 > K$  implies*

$$\|S_k(t) - S_\infty(t)\| \leq \delta e^{-\beta t}, \quad \forall t \geq 0.$$

**Proof:** Take  $g \in L^2$  such that  $\|g\| \leq 1$ , then

$$\begin{aligned} & \| [S_k(t) - S_\infty(t)]g \| \\ & \leq \| [S_k(t) - S_\infty(t)](g - s^2 R_\infty(s)^2 g) \| + s^2 \| [S_k(t) - S_\infty(t)] R_\infty^2 g \| \equiv \tilde{I} + \tilde{II}. \end{aligned}$$

Now from Lemma 5.1, part 1 there exists  $s_0 > 0$  so that  $s > s_0$  implies

$$\|g - s^2 R_\infty(s)^2 g\| \leq \frac{\delta}{4}.$$

Thus by Lemma 5.1 part 2 we have

$$\tilde{I} \leq (\|S_k(t)\| + \|S_\infty(t)\|) \frac{\delta}{4} \leq e^{-\epsilon \lambda_1(k)t} \frac{\delta}{2}.$$

Now by Lemma 5.1 part 4

$$\begin{aligned} \tilde{II} & \leq s^2(t+2)e^{-\epsilon \lambda_1(k)t} \|R_k(s) - R_\infty(s)\| \\ & \leq s^2 C e^{-\epsilon \lambda_1(k)t/2} \|R_k(s) - R_\infty(s)\| \leq \frac{\delta}{2} \end{aligned}$$

where

$$C = \max_{t \geq 0} (t+2)e^{-\epsilon \lambda_1(k)t/2} \leq \frac{\max(2, 2e^{\epsilon \lambda_1(k)-1})}{\epsilon \lambda_1(k)}$$

and where for  $s$  fixed as above we have chosen  $K$  so that for  $k_0, k_1 > K$  we have

$$\|R_k(s) - R_\infty(s)\| \leq \frac{\delta}{2s^2 C}.$$

$\square$

**Lemma 5.3.** *For every  $\psi \in L^2(0, 1)$ , the bounded operators*

$$(5.2) \quad B_k = A_{k,j}^{1/2} A_{\infty,j}^{-1/2}$$

*converge to the identity in  $L^2$  in the strong operator topology. Further if  $C \subset L^2(0, 1)$  is a relatively compact set in  $L^2$ , then*

$$\sup_{\psi \in C} \|[B_k - I]\psi\| \rightarrow 0, \quad k_0, k_1 \rightarrow \infty.$$

**Proof:**  $B_k$  is bounded by the closed graph theorem and the first part of the proof follows from norm resolvent convergence. See, for example, the proof of Theorem 3.13, page 459 of [11]. The second part is a simple general fact which can be found for example in [9], Theorem 3.2 page 124.  $\square$

**Lemma 5.4.** For  $\phi \in \mathcal{H}^1$  and  $\delta > 0$  there exists a  $K$  such that for  $k_0, k_1 > K$

$$I = C_{\beta}^{1/2} \|A_{k,\beta}^{1/2} [S_{k,\beta}(t) - S_{\infty,\beta}(t)] \phi\| \leq \delta$$

for all  $t \geq 0$ .

**Proof:** For each fixed  $t > 0$  a simple Banach algebra argument can be employed to establish this result (cf. [13] Theorem VIII.20, page 286 ). But to obtain the result uniformly in  $t \geq 0$  a bit more work is required.

$$\begin{aligned} \|A_{k,\beta}^{1/2} [S_{k,\beta}(t) - S_{\infty,\beta}(t)] \phi\| &\leq \| [A_{k,\beta}^{1/2} S_{k,\beta}(t) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t)] \phi \| \\ &+ \| [A_{k,\beta}^{1/2} - A_{\infty,\beta}^{1/2}] S_{\infty,\beta}(t) \phi \| = \| [S_{k,\beta}(t) A_{k,\beta}^{1/2} A_{\infty,\beta}^{-1/2} - S_{\infty,\beta}(t) A_{\infty,\beta}^{1/2}] \phi \| \\ &+ \| [A_{k,\beta}^{1/2} A_{\infty,\beta}^{-1/2} - I] A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t) \phi \| \leq \| [S_{k,\beta}(t) [A_{k,\beta}^{1/2} A_{\infty,\beta}^{-1/2} - I] A_{\infty,\beta}^{1/2} \phi \| \\ &+ \| [S_{k,\beta}(t) - S_{\infty,\beta}(t)] A_{\infty,\beta}^{1/2} \phi \| \\ &+ \| [A_{k,\beta}^{1/2} A_{\infty,\beta}^{-1/2} - I] A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t) \phi \| \\ &= I + \tilde{I} + \tilde{I}I. \end{aligned}$$

Now define  $\psi = A_{\infty,\beta}^{1/2} \phi$ . For the first term we have

$$(5.3) \quad \tilde{I} \leq e^{\beta t} \| [A_{k,\beta}^{1/2} A_{\infty,\beta}^{-1/2} - I] \psi \| \leq C_{\beta}^{-1/2} \delta / 3$$

for  $k_0, k_1 > K_1$ . For the second term

$$(5.4) \quad \tilde{I}I := \| [S_{k,\beta}(t) - S_{\infty,\beta}(t)] \psi \| \leq C_{\beta}^{-1/2} \delta / 3$$

for  $k_0, k_1 > K_2$  by Lemma 5.2. And finally for the third term we note that the set

$$\{S_{\infty,\beta}(t)\psi\}_{t \geq 0}$$

is a relatively compact set in  $L^2(0, 1)$  for  $\psi \in L^2(0, 1)$ . Hence by Lemma 5.3, we have

$$(5.5) \quad \tilde{I}I \leq \sup_{t \geq 0} \| [A_{k,\beta}^{1/2} A_{\infty,\beta}^{-1/2} - I] S_{\infty,\beta}(t) \psi \| \leq C_{\beta}^{-1/2} \delta / 3.$$

Combining (5.3), (5.4) and (5.5) the result follows.  $\square$

We have left to consider the two remaining terms  $I\tilde{I}$  and  $\tilde{I}I$  in (5.1).

$$IIb = \left\| \int_0^t [A_{k,\beta}^{1/2} S_{k,\beta}(t-\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t-\tau)] e^{t\beta\tau} F(z_\infty(\tau)) d\tau \right\|,$$

$$C_\beta^{1/2} III = \left\| \int_0^t [A_{k,\beta}^{1/2} - A_{\infty,\beta}^{1/2}] S_{\infty,\beta}(t-\tau) e^{t\beta\tau} F(z_\infty(\tau)) d\tau \right\|.$$

**Lemma 5.5.** *For any  $\delta > 0$  there exists a  $K > 0$  so that for  $k_0, k_1 > K$  we have*

$$C_\beta^{1/2} III \leq \delta$$

for all  $t \geq 0$ .

**Proof:** First note that with  $B_k$  defined in (5.2)

$$C_\beta^{1/2} III = \left\| [B_k - I] \int_0^t A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t-\tau) e^{t\beta\tau} F(z_\infty(\tau)) d\tau \right\|.$$

Now we show that the set

$$\mathcal{S} = \bigcup_{t \geq 0} \left\{ \int_0^t A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t-\tau) e^{t\beta\tau} F(z_\infty(\tau)) d\tau \right\}$$

is a relatively compact set in  $L^2(0, 1)$  and hence the result will follow from Lemma 5.3. Take  $\alpha = 1 + \gamma$  where  $(1 + \gamma)/2 < 1$ . Now recall from (4.4) and (4.1) that

$$e^{t\beta\tau} \|F(z_\infty(\tau))\| \leq e^{-\epsilon\beta\tau} \|\phi\|^2.$$

We proceed to show that the set  $\mathcal{S}$  is bounded in  $\mathcal{H}^\gamma$  and hence it is a relatively compact set in  $L^2(0, 1)$ . We have

$$\begin{aligned} & \left\| \int_0^t A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t-\tau) e^{t\beta\tau} F(z_\infty(\tau)) d\tau \right\|_\gamma \\ &= \left\| \int_0^t A_{\infty,\beta}^\alpha S_{\infty,\beta}(t-\tau) e^{t\beta\tau} F(z_\infty(\tau)) d\tau \right\| \\ &\leq \epsilon \|\phi\|^2 \int_0^t \|A_{\infty,\beta}^{\alpha/2} S_{\infty,\beta}(t-\tau)\| d\tau \\ &\leq \tilde{C} \int_0^\infty \left[ \left( \frac{\alpha}{2\epsilon s} \right)^{\alpha/2} + \lambda_1(k)^{\alpha/2} \right] e^{-\epsilon(\lambda_1(k) - \beta)s} ds \\ &\leq \tilde{\tilde{C}} < \infty. \end{aligned}$$

□

For the term  $Ib$  we extend a result from Simon and Reed (cf. [13] Theorem VIII.20, page 286 ).

**Lemma 5.6.** *For any  $\delta > 0$  there exists a  $K > 0$  so that for  $k_0, k_1 > K$  we have*

$$Ib \leq \delta$$

for all  $t \geq 0$ .

**Proof:** First we need to reduce our calculations to a compact time interval. Thus consider

$$\begin{aligned} & \left\| \int_0^t [A_{k,\beta}^{1/2} S_{k,\beta}(t-\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t-\tau)] e^{c\beta\tau} F(z_\infty(\tau)) d\tau \right\| \\ &= \left\| \int_0^t [A_{k,\beta}^{1/2} S_{k,\beta}(\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(\tau)] e^{c\beta(t-\tau)} F(z_\infty(t-\tau)) d\tau \right\| \\ &\leq c\|\phi\|^2 \int_0^t \|[A_{k,\beta}^{1/2} S_{k,\beta}(\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(\tau)]\| d\tau \\ &= c\|\phi\|^2 \int_0^{t_0} \|[A_{k,\beta}^{1/2} S_{k,\beta}(\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(\tau)]\| d\tau \\ &\quad + c\|\phi\|^2 \int_{t_0}^{t_1} \|[A_{k,\beta}^{1/2} S_{k,\beta}(\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(\tau)]\| d\tau \\ &\quad + c\|\phi\|^2 \int_{t_1}^\infty \|[A_{k,\beta}^{1/2} S_{k,\beta}(\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(\tau)]\| d\tau. \end{aligned}$$

For the first term after the last equality above we have

$$\begin{aligned} & \int_0^{t_0} \|[A_{k,\beta}^{1/2} S_{k,\beta}(\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(\tau)]\| d\tau \\ &\leq \int_0^{t_0} \left( \|[A_{k,\beta}^{1/2} S_{k,\beta}(\tau)]\| + \|[A_{\infty,\beta}^{1/2} S_{\infty,\beta}(\tau)]\| \right) d\tau \\ &\leq \tilde{C} \int_0^{t_0} \frac{e^{-c(\lambda_1(k) - \beta)s}}{\sqrt{s}} ds \\ &\leq \tilde{C} \int_0^{t_0} s^{-1/2} ds \leq \tilde{C} \sqrt{t_0} \leq \delta/3 \end{aligned}$$

for  $t_0$  small enough.



For the last term above

$$\begin{aligned}
 & \int_{t_1}^{\infty} \| [A_{k,\beta}^{1/2} S_{k,\beta}(\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(\tau)] \| d\tau \\
 & \leq \int_{t_1}^{\infty} \left( \| A_{k,\beta}^{1/2} S_{k,\beta}(\tau) \| + \| A_{\infty,\beta}^{1/2} S_{\infty,\beta}(\tau) \| \right) d\tau \\
 & \leq \tilde{C} \int_{t_1}^{\infty} \frac{e^{-\epsilon(\lambda_1(k) - \beta)s}}{\sqrt{s}} ds \\
 & \leq \frac{\tilde{C}}{\sqrt{t_1}} \int_0^{\infty} e^{-\epsilon(\lambda_1(k) - \beta)s} ds \leq \frac{\tilde{C}}{\sqrt{t_1} \epsilon (\lambda_1(k) - \beta)} \leq \delta/3
 \end{aligned}$$

for  $t_1$  large enough.

Finally, with  $t_0$  and  $t_1$  chosen above, we need only consider the integral over the fixed region  $[t_0, t_1]$  which is compact. Thus we consider the term

$$\int_{t_0}^{t_1} \| [A_{k,\beta}^{1/2} S_{k,\beta}(\tau) - A_{\infty,\beta}^{1/2} S_{\infty,\beta}(\tau)] \| d\tau.$$

Let

$$g_{\beta}(\lambda, t) = \begin{cases} (\lambda - \beta)^{1/2} e^{-\epsilon(\lambda - \beta)t}, & \lambda > \beta, \\ 0, & \lambda \leq \beta. \end{cases}$$

The function  $g_{\beta}$  is a uniformly continuous function on  $\mathcal{R} \times [t_0, t_1]$  and vanishes at infinity in  $\lambda$ . Furthermore by the spectral theorem

$$g_{\beta}(A_k, t) = A_{k,\beta}^{1/2} S_{k,\beta}(t), \quad g_{\beta}(A_{\infty}, t) = A_{\infty,\beta}^{1/2} S_{\infty,\beta}(t),$$

and since  $[t_0, t_1]$  is compact we can find  $\{s_j\}_{j=1}^N$  and disjoint intervals  $\{I_j\}$  such that

$$[t_0, t_1] = \bigcup_{j=1}^N \bar{I}_j$$

and

$$\sup_{\lambda \in \mathcal{R}, t \in I_j} |g_{\beta}(\lambda, t) - g_{\beta}(\lambda, s_j)| \leq \frac{\delta}{9(t_1 - t_0)}$$

for  $j = 1, \dots, N$  which implies by the spectral theorem

$$\sup_{\lambda \in \mathcal{R}, t \in I_j} \| A_{k,\beta}^{1/2} S_{k,\beta}(t) - A_{k,\beta}^{1/2} S_{k,\beta}(s_j) \| \leq \frac{\delta}{9(t_1 - t_0)}$$

for  $j = 1, \dots, N$ . A similar expression holds for  $k = \infty$ .

Finally, and once again as a consequence of the spectral theorem (cf. [13] Theorem VIII.20, page 286) and the norm resolvent convergence as  $k_0, k_1$  go to infinity, that there exists a  $K > 0$  so that for  $k_0, k_1 > K$  and for  $j = 1, \dots, N$

$$\sup_{t \in I_j} \|A_{k,j}^{1/2} S_{k,j}(s_j) - A_{\infty,j}^{1/2} S_{\infty,j}(s_j)\| \leq \frac{\delta}{9(t_1 - t_0)}.$$

Combining these results we can compute

$$\begin{aligned} & \int_{t_0}^{t_1} \| [A_{k,j}^{1/2} S_{k,j}(\tau) - A_{\infty,j}^{1/2} S_{\infty,j}(\tau)] \| d\tau \\ &= \sum_{j=1}^N \int_{I_j} \| [A_{k,j}^{1/2} S_{k,j}(\tau) - A_{\infty,j}^{1/2} S_{\infty,j}(\tau)] \| d\tau \\ &\leq \sum_{j=1}^N \left\{ \int_{I_j} \| [A_{k,j}^{1/2} S_{k,j}(\tau) - A_{k,j}^{1/2} S_{k,j}(s_j)] \| d\tau \right. \\ &\quad + \int_{I_j} \| [A_{k,j}^{1/2} S_{k,j}(s_j) - A_{\infty,j}^{1/2} S_{\infty,j}(s_j)] \| d\tau \\ &\quad \left. + \int_{I_j} \| [A_{\infty,j}^{1/2} S_{\infty,j}(\tau) - A_{\infty,j}^{1/2} S_{\infty,j}(s_j)] \| d\tau \right\} \\ &\leq \frac{\delta}{9(t_1 - t_0)} [3(t_1 - t_0)] \leq \frac{\delta}{3}. \end{aligned}$$

□

## References

1. C.I. Byrnes, D.S. Gilliam, "Boundary feedback design for nonlinear distributed parameter systems," *Proc. of IEEE Conf. on Dec. and Control*, Britton, England 1991.
2. C.I. Byrnes and D.S. Gilliam, "Stability of certain distributed parameter systems by low dimensional controllers: a root locus approach," *Proceedings 29th IEEE International Conference on Decision and Control*.
3. C.I. Byrnes and D.S. Gilliam, "Boundary feedback stabilization of a viscous Burgers' equation," to appear in *Proceedings Third Conference on Computation and Control*, Bozeman, Montana.
4. C.I. Byrnes and D.S. Gilliam, J. He, "Root Locus and Boundary Feedback design for a Class of Distributed Parameter Systems," to appear in *SIAM J. Control and Opt.*
5. C.I. Byrnes, D.S. Gilliam and V.I. Shubov, "Boundary control for a viscous burgers' equation," *Identification and control in systems governed by partial differential equations* H.T. Banks, R.H. Fabiano and K. Ito (Eds.), SIAM, 1993, 171-185.
6. C.I. Byrnes, D.S. Gilliam and V.I. Shubov, "High gain limits of trajectories and attractors for a controlled viscous Burgers' equation," Submitted to *J. Math. Sys. Est. Control*.
7. C.I. Byrnes, D.S. Gilliam and V.I. Shubov, "Zero and pole dynamics of a nonlinear distributed parameter system," Preprint.
8. J. Carr, "Applications of the center manifold theorem," Springer-Verlag, 1980.

9. F. Chatelin, "Spectral Approximation of Linear Operators," Academic Press, Series in Computer Science and Applied Mathematics, 1983.
10. D. Henry, "Geometric Theory of Semilinear Parabolic Equations," Springer-Verlag, *Lect. Notes in Math.*, Vol 840, 1981.
11. T. Kato, "Perturbation theory for linear operators," Springer-Verlag, New York, (1966).
12. A. Pazy, "Semigroups of Linear Operators and Applications to Partial Differential Equations," Springer-Verlag, 1983.
13. B. Simon and M. Reed, "Methods of Modern Mathematical Physics, Volume 1," Academic Press, New York and London, 1972.

Christopher I. Byrnes  
Department of Systems Science and Mathematics  
Washington University  
St. Louis, MO 63130, USA

David S. Gilliam  
Department of Mathematics  
Texas Tech University  
Lubbock, TX 79409, USA

Victor I. Shubov  
Department of Mathematics  
Texas Tech University  
Lubbock, TX 79409, USA

## OPTIMALITY CONDITIONS FOR BOUNDARY CONTROL PROBLEMS OF PARABOLIC TYPE

PIERMARCO CANNARSA AND MARIA ELISABETTA TESSITORE

Dipartimento di Matematica  
Università di Roma "Tor Vergata"

Dipartimento di Matematica  
Università di Roma "La Sapienza"

**ABSTRACT.** This paper is devoted to the study of finite horizon optimal control problems with boundary control. We prove a sufficient condition for optimality of trajectory-control pairs, using a non-smooth analysis approach. We formulate this condition in terms of an Hamiltonian system for which we show an existence and uniqueness result.

**1991 Mathematics Subject Classification.** 49K20, 49K27

**Key words and phrases.** Optimality conditions, boundary control, parabolic equations, Neumann boundary conditions, analytic semigroups, backward uniqueness.

### 1. Introduction

This paper is concerned with the problem of minimizing

$$(1.1) \quad \int_{t_0}^T L(s, x(s; t_0, x_0, u), u(s)) ds + \phi(x(T; t_0, x_0, u)),$$

overall measurable functions  $u : [t_0, T] \rightarrow U$  (such a function is usually called a *control*). Here  $(t_0, x_0) \in [0, T] \times X$ , and  $x(s; t_0, x_0, u)$  is the mild solution of the infinite dimensional controlled system

$$(1.2) \quad \begin{cases} x'(t) = Ax(t) + F(x(t)) - ANu(t) \\ x(t_0) = x_0. \end{cases}$$

The state space  $X$  and the control space  $U$  are real Hilbert spaces and  $A$  is the infinitesimal generator of an analytic semigroup on  $X$ . Moreover,  $F : X \rightarrow X$  is Lipschitz continuous and  $N$  is the so called Neumann map.

Equation (1.2) is the abstract version of a parabolic equation that is controlled through a Neumann type boundary condition. Further details on this connection are recalled in section 2, where we also list all the assumptions on the

various data. Here, we just wish to mention the fact that the running cost  $L$  is allowed to be unbounded, provided that a coercivity condition of the form

$$\exists \lambda_0 > 0, \lambda_1 \in \mathbb{R} : L(t, x, u) \geq \lambda_0 |u|^2 + \lambda_1, \forall t \in [0, T], u \in U.$$

is satisfied.

The literature on boundary control is huge. Clearly, the most investigated case is the Linear Quadratic problem, see for instance [16], [3], [15]. Among nonlinear boundary control problems one of the first examples to be studied was the convex case, see [1], [13], that requires a linear state equation and a convex payoff. As for general nonlinear boundary control problems, most of the results that are available in the literature are concerned with necessary optimality conditions, see e.g. [9], [10] and [20]. The Dynamic Programming approach to nonlinear boundary control problems is more recent and uses viscosity solutions, see [6] and [7].

In this paper we focus our attention on both necessary and sufficient conditions for optimality. Necessary conditions will be stated both in the well known traditional form of the Pontryagin Maximum Principle (see Theorem (3.1)) and in the Hamiltonian formulation

$$(1.3) \quad \begin{cases} x'(t) = Ax(t) + F(x(t)) - (-A)^j D_p H(t, x(t), (-A)^j p(t)) \\ -p'(t) = Ap(t) + [D_x F(x(t))]^* p(t) - D_x H(t, x(t), (-A)^j p(t)) \end{cases}$$

where  $H$  is the Hamiltonian associated with  $L$ . As usual, system (1.3) is assigned the initial-terminal condition

$$(1.4) \quad \begin{cases} x(t_0) = x_0 \\ p(T) = -D\varphi(x(T)) \end{cases}.$$

As it is done in [4] for distributed control problems, the formulation of the Pontryagin Maximum Principle is complemented with co-state inclusions of the type derived in [8] for finite dimensional control problems. Roughly speaking, such inclusions establish a connection between the co-state associated with some fixed optimal pair  $\{x^*, u^*\}$ , and some generalized gradient of the value function along the optimal trajectory  $x^*$ . We recall that the value function of problem (1.1), (1.2) is defined as

$$(1.5) \quad V(t_0, x_0) = \inf_{u \in U} \left\{ \int_{t_0}^T L(s, x(s; t_0, x_0, u), u(s)) ds + \phi(x(T; t_0, x_0, u)) \right\}.$$

The main object of this paper, however, is the generalization of a sufficient optimality condition obtained in [4] for distributed control problems. Such a result is based on the idea of replacing the initial-terminal condition (1.4) with an initial condition of the form

$$(1.6) \quad \begin{cases} x(t_0) = x_0 \\ p(t_0) = p_0 \end{cases}$$

where  $-p_0$  is a cluster point for the  $x$  gradient of the value function  $V$  at  $(t_0, x_0)$ . More precisely, extending the result of [4] to boundary control, we prove that, for any such  $p_0$ , problem (1.3)–(1.6) has a solution whose  $x$  component is an optimal trajectory of (1.1)–(1.2). Other kinds of sufficient conditions for boundary control problems are discussed in [12].

Moreover, in this paper, we prove that the solution of system (1.3)–(1.6) is *unique*. This uniqueness result seems to be new even for distributed control problems. Since in system (1.3) one equation is forward and the other one is backward, a natural approach to the problem of uniqueness is to use backward uniqueness techniques. A detailed discussion of these techniques is contained in [14]. In this paper, however, we adapt (as in [5]) the method of [17].

We conclude with an outline of the paper. In §2 we recall basic material on boundary control problems, including some properties of the value function of problem (1.1)–(1.2). In §3 we state necessary conditions for optimality through the Pontryagin Maximum Principle (Theorem 3.1). We then formulate its Hamiltonian version and we show an existence and uniqueness result for this system, which is also a sufficient condition for optimality (Theorems 3.3 and 3.4).

## 2. Preliminaries

Let  $T > 0$ ,  $t_0 \in [0, T]$  and  $\Omega \subset \mathbb{R}^n$  be open and bounded with smooth boundary. Consider the following boundary value problem with Neumann conditions

$$(2.1) \quad \begin{cases} \frac{\partial x}{\partial t}(t, \xi) = \Delta_{\xi} x(t, \xi) + f(x(t, \xi)) & \text{in } (t_0, T) \times \Omega \\ x(t_0, \xi) = x_0(\xi) & \text{on } \Omega \\ \frac{\partial x}{\partial \nu}(t, \xi) = u(t, \xi) & \text{on } (t_0, T) \times \Omega \end{cases}$$

where  $x_0 \in L^2(\Omega)$ ,  $u \in L^2(t_0, T; L^2(\partial\Omega))$ , and  $f: \mathbb{R} \rightarrow \mathbb{R}$  is Lipschitz continuous.

Given continuous functions  $L: [0, T] \times L^2(\Omega) \times L^2(\partial\Omega) \rightarrow \mathbb{R}$  and  $\phi: L^2(\Omega) \rightarrow \mathbb{R}$ , let us consider the problem of minimizing the functional

$$(2.2) \quad J(t_0, x_0; u) = \int_{t_0}^T L(t, x(t, \cdot), u(t, \cdot)) dt + \phi(x(T, \cdot))$$

overall controls  $u \in L^2(t_0, T; L^2(\partial\Omega))$ , where  $x$  is the solution of (2.1).

Problem (2.1)–(2.2) may be rewritten in abstract form as follows.

Let  $X = L^2(\Omega)$ ,  $U = L^2(\partial\Omega)$  and define the unbounded operator  $A$  in  $X$  by

$$D(A) = \left\{ x \in H^2(\Omega) : \frac{\partial x}{\partial \nu} = 0 \right\}$$

$$Ax = \Delta x - x.$$

Next, we define the Neumann map  $N : U \rightarrow X$  as (see [9])

$$Nu = x \Leftrightarrow \begin{cases} \Delta x = x & \text{in } \Omega \\ \frac{\partial x}{\partial \nu} = u & \text{on } \partial\Omega \end{cases}$$

Formally, equation (2.1) may be written as

$$(2.3) \quad \begin{cases} x'(t) = Ax(t) + F(x(t)) - (Nu(t)) \\ x(t_0) = x_0 \end{cases}$$

where

$$F(x)(\xi) = f(x(\xi)) + x(\xi).$$

The right hand side of the equation (2.3) is not well defined as the range of  $N$  is not contained in  $D(A)$ . However, we note that  $N$  has some regularizing effect. Indeed,  $N : L^2(\partial\Omega) \rightarrow H^{\frac{3}{2}}(\Omega)$ , which may be expressed in abstract form using the fractional powers of  $-A$ . In fact,

$$D((-A)^\theta) = \begin{cases} H^{2\theta}(\Omega) & \text{if } 0 \leq \theta < \frac{3}{4} \\ \left\{ x \in H^{2\theta}(\Omega) : \frac{\partial x}{\partial \nu} = 0 \right\} & \text{if } \frac{3}{4} < \theta \leq 1. \end{cases}$$

The fact that operator  $A$  is the generator of an analytic semigroup in  $X$  implies that for every  $\theta \in [0, 1]$  there exists a constant  $M_\theta > 0$  such that

$$(2.4) \quad |(-A)^\theta e^{tA}x| \leq \frac{M_\theta}{t^\theta} |x|, \quad \forall t > 0, \forall x \in X.$$

Moreover let  $\alpha \in ]0, \frac{1}{2}[$ . Then, a well known interpolation inequality, see e.g. [18], states that for every  $\sigma > 0$  there exists  $C'_\sigma > 0$  such that

$$(2.5) \quad |(-A)^\alpha x| \leq \sigma |(-A)^{\frac{1}{2}}x| + C'_\sigma |x|, \quad \forall x \in D((-A)^{\frac{1}{2}}).$$

By classical regularity results,  $N : U \rightarrow D((-A)^\beta)$  for all  $0 < \beta < \frac{3}{4}$ . Therefore, equation (2.3) can be written as

$$(2.6) \quad \begin{cases} x'(t) = Ax(t) + F(x(t)) + (-A)^\beta N_\beta u(t) \\ x(t_0) = x_0 \end{cases}$$

where  $N_\beta = (-A)^{1-\beta}N$ . Clearly, equation (2.6) has to be understood in mild form, that is

$$(2.7) \quad x(t) = e^{(t-t_0)A}x_0 + \int_{t_0}^t e^{(t-s)A}F(x(s))ds + (-A)^\beta \int_{t_0}^t e^{(t-s)A}N_\beta u(s)ds \quad t_0 \leq t \leq T.$$

It is well known that (2.7) has a unique solution

$$x(\cdot; t_0, x_0, u) \in C([t_0, T]; X)$$

under the following assumptions:

$$(2.8) \quad \begin{cases} (i) & A : D(A) \subset X \rightarrow X \text{ is a densely defined closed linear operator;} \\ (ii) & \text{the inclusion } D(A) \subset X \text{ is compact and } A = A^*; \\ (iii) & \exists \omega > 0 \text{ such that } \langle Ax, x \rangle \leq -\omega |x|^2, \forall x \in D(A); \\ (iv) & F : X \rightarrow X \text{ is Lipschitz continuous;} \\ (v) & \beta \in ]\frac{1}{4}, \frac{1}{2}[. \end{cases}$$

The value function of problem (2.1)–(2.2) is defined as

$$(2.9) \quad V(t_0, x_0) = \inf_{u \in L^2(t_0, T; U)} \left\{ \int_{t_0}^T L(s, x(s; t_0, x_0, u), u(s)) ds + \phi(x(T; t_0, x_0, u)) \right\}$$

where  $L : [0, T] \times X \times U \rightarrow \mathbb{R}$  and  $\phi : X \rightarrow \mathbb{R}$  are assumed to satisfy the following  $\forall R > 0$  and some constant  $C_R > 0$ :

$$(2.10) \quad \begin{cases} (i) & L \in C([0, T] \times X \times U) \\ (ii) & |L(t, x, u) - L(t, y, u)| \\ & \leq C_R |x - y|, \forall t \in [0, T], u \in U, |x|, |y| \leq R \\ (iii) & L(t, x, \cdot) \text{ is strictly convex} \\ (iv) & \exists \lambda_0 > 0, \lambda_1 \in \mathbb{R} : L(t, x, u) \geq \lambda_0 |u|^2 + \lambda_1, \forall t \in [0, T], u \in U \\ (v) & |L(t, x, u) - L(t, x, v)| \\ & \leq C_R (1 + |u| + |v|) |u - v|, \forall t \in [0, T], u, v \in U, |x| \leq R \\ (vi) & |\phi(x) - \phi(y)| \leq C_R |x - y|, \forall |x|, |y| \leq R. \end{cases}$$

A control  $u^* \in L^2(t_0, T; U)$  at which the infimum in (2.9) is attained, is said to be *optimal*. Equivalently,  $u^*$  is optimal if

$$V(t_0, x_0) = \int_{t_0}^T L(s, x(s; t_0, x_0, u^*), u^*(s)) ds + \phi(x(T; t_0, x_0, u^*)).$$

The following Lemma will be useful in the sequel to prove the Pontryagin Maximum Principle. We denote by  $J$  the payoff functional associated to problem (2.9), i.e.

$$(2.11) \quad J(t_0, x_0; u) = \int_{t_0}^T L(t, x(t; t_0, x_0, u), u(t)) dt + \phi(x(T; t_0, x_0, u)).$$



**Lemma 2.1.** Assume (2.8), (2.10). Then, for any  $R > 0$  there exists a constant  $M_R > 0$  such that, for any  $x_0 \in X$ , with  $|x_0| \leq R$ , and any control  $u \in L^2(t_0, T; U)$ , there exists  $\bar{u} \in L^2(t_0, T; U)$  satisfying

$$(i) \quad J(t_0, x_0; \bar{u}) \leq J(t_0, x_0; u)$$

$$(ii) \quad |\bar{u}(t)| \leq \frac{M_R}{(T-t)^\beta}, \forall t \in [t_0, T].$$

**Proof** In the following, we will assume that hypothesis (2.10) holds true with  $C_R = C$  independent of  $R$ . Then, the constant  $M_R$  in (ii) above is also independent of  $R$ , as we will show below. The proof of the general case is similar and is left to the reader.

Let  $u \in L^2(t_0, T; U)$  and define, for any  $n \in \mathbb{N}$ ,

$$I_n = \left\{ t \in [t_0, T] : |u(t)| > \frac{n}{(T-t)^\beta} \right\}$$

and

$$u_n(t) = \begin{cases} u(t) & \text{if } t \notin I_n \\ 0 & \text{if } t \in I_n. \end{cases}$$

Moreover, let us set

$$x(t) = x(t; t_0, x_0, u), \quad x_n(t) = x(t; t_0, x_0, u_n)$$

Then, denoting by  $|I_n|$  the Lebesgue measure of  $I_n$ , we have

$$\begin{aligned} J(t_0, x_0; u_n) &= J(t_0, x_0; u) + \int_{t_0}^T [L(t, x_n(t), u_n(t)) - L(t, x_n(t), u(t))] dt \\ &\quad + \int_{t_0}^T [L(t, x_n(t), u(t)) - L(t, x(t), u(t))] dt + [\phi(x_n(T)) - \phi(x(T))] \\ (2.12) \quad &\leq J(t_0, x_0; u) + |\lambda_1| |I_n| + \lambda_0 \int_{I_n} |u(r)|^2 dr \\ &\quad + C \int_{t_0}^T |x_n(t) - x(t)| dt + C |x_n(T) - x(T)|. \end{aligned}$$

Now, recalling (2.7) and (2.4),

$$(2.13) \quad |x_n(s) - x(s)| \leq C_1 \int_{t_0}^s |x_n(r) - x(r)| dr + C_1 \int_{t_0}^s \frac{|u(r)|}{(s-r)^\beta} \chi_{I_n}(r) dr$$

where  $\chi_{I_n}$  denotes the characteristic function of the set  $I_n$ . Let

$$\eta(t) = \int_{t_0}^t |x_n(s) - x(s)| ds.$$

Then, integrating (2.13),

$$\begin{aligned} \eta(t) &\leq C_1 \int_{t_0}^t \eta(s) ds + C_1 \int_{t_0}^t ds \int_{t_0}^s \frac{|u(r)|}{(s-r)^\beta} \chi_{I_n}(r) dr \\ &= C_1 \int_{t_0}^t \eta(s) ds + \frac{C_1}{1-\beta} \int_{t_0}^t |u(r)|(t-r)^{1-\beta} \chi_{I_n}(r) dr \\ &\leq C_1 \int_{t_0}^t \eta(s) ds + \frac{C_1 T^{1-\beta}}{1-\beta} \int_{I_n} |u(r)| dr. \end{aligned}$$

Hence, by Gronwall's Inequality,

$$\eta(t) \leq e^{C_1 T} \frac{C_1 T^{1-\beta}}{1-\beta} \int_{I_n} |u(r)| dr = C_2 \int_{I_n} |u(r)| dr.$$

From (2.13) and the estimate above, it follows that

$$(2.14) \quad |x_n(s) - x(s)| \leq C_3 \int_{I_n} |u(r)| dr + C_1 \int_{t_0}^s \frac{|u(r)|}{(s-r)^\beta} \chi_{I_n}(r) dr.$$

Using (2.14) in (2.12) we obtain, after some tedious computations,

$$\begin{aligned} (2.15) \quad &J(t_0, x_0; u_n) - J(t_0, x_0; u) \\ &\leq |\lambda_1| |I_n| - \lambda_0 \int_{I_n} |u(r)|^2 dr + C_4 \int_{I_n} |u(r)| dr + C_5 \int_{I_n} \frac{|u(r)|}{(T-r)^\beta} dr \end{aligned}$$

Finally, we claim that the right hand side of (2.15) is negative for sufficiently large  $n$ , which will yield the conclusion of the Lemma. Indeed,

$$|\lambda_1| |I_n| - \frac{1}{3} \lambda_0 \int_{I_n} |u(r)|^2 dr - |\lambda_1| |I_n| - \frac{n^2 \lambda_0}{3T^{2\beta}} |I_n| < 0$$

provided  $n$  is large enough, say  $n \geq n_1$ . Similarly,

$$C_4 \int_{I_n} |u(r)| dr - \frac{1}{3} \lambda_0 \int_{I_n} |u(r)|^2 dr \leq C_4 \int_{I_n} |u(r)| dr - \frac{n \lambda_0}{3T^{2\beta}} \int_{I_n} |u(r)| dr < 0$$

if  $n \geq n_2$ . Furthermore,

$$C_5 \int_{I_n} \frac{|u(r)|}{(T-r)^\beta} dr - \frac{1}{3} \lambda_0 \int_{I_n} |u(r)|^2 dr \leq C_5 \int_{I_n} \frac{|u(r)|}{(T-r)^\beta} dr - \frac{n\lambda_0}{3} \int_{I_n} \frac{|u(r)|}{(T-r)^\beta} dr < 0$$

if  $n \geq n_3$ . The claim follows and the proof is complete. ■

We now recall the definitions of some generalized gradients that will be used in the sequel. Let  $D$  be an open subset of  $X$ . The superdifferential of a function  $v: D \rightarrow \mathbb{R}$  at a point  $x_0 \in D$  is the (possibly empty) set

$$(2.16) \quad D^+ v(x_0) = \left\{ p \in X : \limsup_{x \rightarrow x_0} \frac{v(x) - v(x_0) - \langle p, x - x_0 \rangle}{|x - x_0|} \leq 0 \right\}.$$

We denote by  $D^* v(x_0)$  the set of all vectors  $p \in X$  for which there exists a sequence  $\{x_n\}$  of points of  $D$  such that

$$(2.17) \quad \begin{cases} (i) & x_n \rightarrow x_0 \text{ as } n \rightarrow +\infty \\ (ii) & v \text{ is Fréchet differentiable at } x_n, \forall n \\ (iii) & Dv(x_n) \rightarrow p \text{ as } n \rightarrow +\infty \end{cases}$$

If  $v$  is Lipschitz in a neighborhood  $D_0$  of  $x_0$ , then  $v$  is Fréchet differentiable on a dense subset of  $D_0$  (see [19]). Hence,  $D^* v(x_0) \neq \emptyset$ .

Next, we recall a Lipschitz regularity result for the value function  $V$  defined in (2.9) proved in [7] (see also [6]).

**Proposition 2.2.** *Assume (2.8), (2.10). Then, the value function  $V$  defined in (2.9) is continuous on  $[0, T] \times X$ . Moreover, for any  $R > \frac{1}{T}$  and  $\theta \in [0, 1[$ , there exists  $C_{R,\theta} > 0$  such that*

$$|V(t, x) - V(t, y)| \leq C_{R,\theta} |(-A)^{-\theta}(x - y)|$$

for all  $t \in [0, T - \frac{1}{R}]$  and all  $x, y \in X$  satisfying  $|x|, |y| \leq R$ . In particular, for all such  $t$  and  $x$ ,

$$(2.18) \quad D_x^+ V(t, x) \subset D((-A)^\theta).$$

### 3. Finite Horizon Problem

In this section we will derive necessary and sufficient conditions for the problem of

$$(3.1) \quad \text{minimizing } J(t_0, x_0; u) \text{ over all controls } u \in L^2(t_0, T; U)$$

where  $J$  is defined in (2.11). In addition to hypotheses (2.8), (2.10), we will assume that

$$(3.2) \quad \begin{cases} (i) & F \text{ is Fréchet differentiable} \\ (ii) & L \text{ is Fréchet differentiable with respect to } x. \end{cases}$$

Notice that the above assumptions and (2.8)–(2.10) imply that  $DF$  and  $D_x L$  are bounded on all bounded subsets of  $X$ . Moreover we assume that for suitable  $c_1, c_2$

$$(3.3) \quad |D_x L(t, x, u)| \leq c_1 L(t, x, u) + c_2$$

Let  $\bar{u} \in L^2(t_0, T; U)$ ,  $\bar{x}(\cdot) = x(\cdot; t_0, x_0, \bar{u})$  and  $p_T \in D^1 \phi(\bar{x}(T))$ .

We recall that the *co state* associated to the triplet  $\{\bar{u}, \bar{x}, p_T\}$  is the mild solution to the problem

$$(3.4) \quad \begin{cases} -p'(t) = Ap(t) + [DF(\bar{x}(t))^* p(t) + D_x L(t, \bar{x}(t), \bar{u}(t))], & t \in [t_0, T] \\ p(T) = -p_T. \end{cases}$$

**Theorem 3.1.** Assume (2.8), (2.10), (3.2) and (3.3). Let  $\{\bar{u}, \bar{x}\}$  be an optimal pair for problem (3.1),  $p_T \in D^1 \phi(\bar{x}(T))$ , and let  $\bar{p}$  be the corresponding co state. Then,

$$(3.5) \quad -\bar{p}(t) \in D_x^1 V(t, \bar{x}(t))$$

for all  $t \in [t_0, T]$  and

$$(3.6) \quad -\langle N_\beta \bar{u}(t), (-A)^j \bar{p}(t) \rangle - L(t, \bar{x}(t), \bar{u}(t)) = H(t, \bar{x}(t), (-A)^j \bar{p}(t))$$

for a.e.  $t \in [t_0, T]$ , where

$$(3.7) \quad H(t, x, p) = \sup_{u \in U} [-\langle N_\beta u, p \rangle - L(t, x, u)].$$

**Proof** In view of (3.5) and of Proposition 2.2 we conclude that  $\bar{p}(t) \in D((-A)^j)$ ,  $\forall t \in [t_0, T]$ . Therefore the left hand side of (3.6) is well defined. From the Dynamic Programming Principle and Lemma (2.1) it follows that for  $|x_0| \leq R$  and for  $t_0 < T_0 < T$

$$(3.8) \quad V(t_0, x_0) = \inf \left\{ \int_{t_0}^{T_0} L(s, x(s; t_0, x_0, u), u(s)) ds + V(T_0, x(T_0; t_0, x_0, u)) : |u| \leq \frac{M_R}{(T - T_0)^j} \right\}$$

Then we can apply the Maximum Principle contained in [4], Theorem 6.1. So we conclude (3.5) and (3.6).

**Remark 3.2.** From hypotheses (2.10)(iii) and (3.2) it follows that  $H$  is differentiable with respect to  $(x, p)$ . So, by (3.6) and (3.7)

$$(3.9) \quad \bar{u}(t) = -N_\beta^* D_p H(t, \bar{x}(t), (-A)^j \bar{p}(t))$$

for a.e.  $t \in [t_0, T]$ . The above equation and (3.5) yield the feedback law

$$(3.10) \quad \bar{u}(t) \in -N_\beta^* D_p H(t, \bar{x}(t), (-A)^j D_x^1 V(t, \bar{x}(t)))$$

for a.e.  $t \in [t_0, T]$ .

The next result may be directly derived following the same reasonings contained in [4], Theorem 5.9.

**Theorem 3.3.** Assume (2.8), (2.10), (3.2) and (3.3). Suppose that  $\alpha$  is Fréchet differentiable and, for all  $R > 0$ ,

$$\|D_x H(t, x, p) - D_x H(t, y, q)\| + \|D_p H(t, x, p) - D_p H(t, y, q)\| \leq C_R \|x - y\| + \|p - q\|$$

for some constant  $C_R > 0$  and all  $x, y, p, q \in X$  satisfying  $\|x\|, \|q\| \leq R$ . Let  $(t_0, x_0) \in [0, T] \times X$  and  $p_0 \in D_t^* V(t_0, x_0)$ . Then, the system

$$(3.11) \quad \begin{cases} x'(t) = Ax(t) + F(x(t)) + (I - A)^{-1} D_p H(t, x(t), (I - A)^{-1} p(t)) \\ p'(t) = Ap(t) + (D_x F(x(t)))^* p(t) - D_x H(t, x(t), (I - A)^{-1} p(t)) \end{cases}$$

with the initial-terminal condition

$$\begin{cases} x(t_0) = x_0 \\ p(T) = -D_x \alpha(F), \end{cases}$$

has a solution  $(x, p)$  such that  $x$  is an optimal trajectory for problem (3.1) corresponding to some control  $u$ . Moreover,  $p$  is the co-state associated to  $u$  and satisfies  $p(t_0) = p_0$ .

The above Theorem gives, in some sense, a sufficient condition for optimality. This condition would be more useful if one could guarantee uniqueness of solutions for (3.11). Uniqueness results for problem (3.11) have been obtained in the linear case, see [11]. In the next Theorem we show an existence and uniqueness result for an Hamiltonian system of kind (3.11) replacing the terminal co-state datum with an initial one. By adapting the same technique of [17] we can derive the following

**Theorem 3.4.** Assume (2.8), (2.10), (3.2) and (3.3). Suppose that

$$\|D_x H(t, x, p) - D_x H(t, y, q)\| + \|D_p H(t, x, p) - D_p H(t, y, q)\| \leq L_H \|x - y\| + \|p - q\|$$

for some constant  $L_H > 0$  and all  $x, y, p, q \in X$ . Let  $(t_0, x_0) \in [0, T] \times X$  and  $p_0 \in D_t^* V(t_0, x_0)$ . Then the Hamiltonian system

$$(3.12) \quad \begin{cases} x'(t) = Ax(t) + (I - A)^{-1} D_p H(t, x(t), (I - A)^{-1} p(t)) - x(t_0) - x_0 \\ p'(t) = Ap(t) - D_x H(t, x(t), (I - A)^{-1} p(t)), \quad p(t_0) = p_0 \end{cases}$$

has a unique solution  $(x, p)$ .

**Proof.** The existence part is a straightforward consequence of the above Theorem. Without loss of generality we set  $t_0 = 0$ . Let us make the following change of variables

$$y(t) = (I - A)^{-1} x(t)$$

Then system (3.12) becomes

$$(3.13) \quad \begin{cases} y'(t) = Ay(t) + D_p H(t, (-A)^{-1}y(t), (-A)^{-1}p(t)), & y(0) = (-A)^{-1}x_0 \\ p'(t) = -Ap(t) + D_x H(t, (-A)^{-1}y(t), (-A)^{-1}p(t)), & p(0) = p_0. \end{cases}$$

Let  $(y_1(t), p_1(t))$  and  $(y_2(t), p_2(t))$  be two distinct solutions to system (3.13) and consider  $q(t) = y_1(t) - y_2(t)$  and  $p(t) = p_1(t) - p_2(t)$ . Then  $q(t)$  and  $p(t)$  satisfy the system

$$(3.14) \quad \begin{cases} q'(t) = Aq(t) + D_p H(t, (-A)^{-1}q(t), (-A)^{-1}p(t)), & q(0) = 0 \\ p'(t) = -Ap(t) + D_x H(t, (-A)^{-1}q(t), (-A)^{-1}p(t)), & p(0) = 0 \end{cases}$$

where

$$\begin{aligned} D_p H(t, (-A)^{-1}q(t), (-A)^{-1}p(t)) \\ D_p H(t, (-A)^{-1}q_1(t), (-A)^{-1}p_1(t)) - D_p H(t, (-A)^{-1}q_2(t), (-A)^{-1}p_2(t)) \end{aligned}$$

and

$$\begin{aligned} D_x H(t, (-A)^{-1}q(t), (-A)^{-1}p(t)) \\ D_x H(t, (-A)^{-1}q_1(t), (-A)^{-1}p_1(t)) - D_x H(t, (-A)^{-1}q_2(t), (-A)^{-1}p_2(t)). \end{aligned}$$

Let  $\theta \in C^1(\mathbb{R})$  be a function such that

$$\theta(t) = \begin{cases} 1 & 0 \leq t \leq \frac{T}{2} \\ 0 & t \geq T \end{cases}$$

$$|\theta'(t)| \leq \frac{1}{T}.$$

We set

$$\tilde{y}(t) = \theta(t)y(t) \text{ and } \tilde{p}(t) = \theta(t)p(t)$$

then  $\tilde{y}(t)$  and  $\tilde{p}(t)$  satisfy the system

$$(3.15) \quad \begin{cases} \tilde{y}'(t) = A\tilde{y}(t) + D_p \tilde{H}(t, (-A)^{-1}\tilde{y}(t), (-A)^{-1}\tilde{p}(t)) + g_y(t), & \tilde{y}(0) = 0 \\ \tilde{p}'(t) = -A\tilde{p}(t) + D_x \tilde{H}(t, (-A)^{-1}\tilde{y}(t), (-A)^{-1}\tilde{p}(t)) + g_p(t), & \tilde{p}(0) = 0 \end{cases}$$

where

$$g_y(t) = \theta'(t)y(t) \text{ and } g_p(t) = \theta'(t)p(t)$$

and

$$D_p \tilde{H}(t, (-A)^{-1}\tilde{y}(t), (-A)^{-1}\tilde{p}(t)) = \theta(t)D_p \tilde{H}(t, (-A)^{-1}\tilde{y}(t), (-A)^{-1}\tilde{p}(t))$$

$$D_x \tilde{H}(t, (-A)^{-1}\tilde{y}(t), (-A)^{-1}\tilde{p}(t)) = \theta(t)D_x \tilde{H}(t, (-A)^{-1}\tilde{y}(t), (-A)^{-1}\tilde{p}(t)).$$

Now we set

$$z(t) = e^{\frac{\kappa(t-T)^2}{2}} \tilde{y}(t) \text{ and } q(t) = e^{\frac{\kappa(t-T)^2}{2}} \tilde{p}(t)$$

then  $z(t)$  and  $q(t)$  satisfy the system

$$(3.16) \quad \begin{cases} z'(t) = Az(t) + k(t-T)z(t) \\ \quad - D_z H(t, (-A)^j z(t), (-A)^j q(t)) + f_z(t), \\ z(0) = z(T) = 0 \\ q'(t) = -Aq(t) + k(t-T)q(t) \\ \quad + D_q H(t, (-A)^j z(t), (-A)^j q(t)) + f_q(t), \\ q(0) = q(T) = 0 \end{cases}$$

where

$$f_z(t) = e^{\frac{k(t-T)^2}{2}} g_x(t) \quad \text{and} \quad f_q(t) = e^{\frac{k(t-T)^2}{2}} g_p(t)$$

and

$$D_z H(t, (-A)^j z(t), (-A)^j q(t)) = e^{\frac{k(t-T)^2}{2}} D_p \bar{H}(t, (-A)^j \bar{y}(t), (-A)^j \bar{p}(t))$$

$$D_q H(t, (-A)^j z(t), (-A)^j q(t)) = e^{\frac{k(t-T)^2}{2}} D_x \bar{H}(t, (-A)^j \bar{y}(t), (-A)^j \bar{p}(t)).$$

Then multiplying the first equation of system (3.16) by  $z'(t)$  and the second equation by  $q'(t)$  we get

$$\begin{aligned} |z'(t)|^2 &= \langle Az(t), z'(t) \rangle + \langle k(t-T)z(t), z'(t) \rangle \\ &\quad - \langle D_z H(t, (-A)^j z(t), (-A)^j q(t)), z'(t) \rangle + \langle f_z(t), z'(t) \rangle \end{aligned}$$

and

$$\begin{aligned} |q'(t)|^2 &= \langle -Aq(t), q'(t) \rangle + \langle k(t-T)q(t), q'(t) \rangle \\ &\quad + \langle D_q H(t, (-A)^j z(t), (-A)^j q(t)), q'(t) \rangle + \langle f_q(t), q'(t) \rangle. \end{aligned}$$

The above equalities can be rewritten as

$$\begin{aligned} |z'(t)|^2 &= \frac{1}{2} \frac{d}{dt} \{ \langle Az(t), z(t) \rangle + k(t-T)|z(t)|^2 \} - \frac{k}{2} |z(t)|^2 \\ &\quad - \langle D_z H(t, (-A)^j z(t), (-A)^j q(t)), z'(t) \rangle + \langle f_z(t), z'(t) \rangle \end{aligned}$$

and

$$\begin{aligned} |q'(t)|^2 &= \frac{1}{2} \frac{d}{dt} \{ \langle -Aq(t), q(t) \rangle + k(t-T)|q(t)|^2 \} - \frac{k}{2} |q(t)|^2 \\ &\quad + \langle D_q H(t, (-A)^j z(t), (-A)^j q(t)), q'(t) \rangle + \langle f_q(t), q'(t) \rangle. \end{aligned}$$

Integrating on  $[0, T]$ , recalling that  $z$  and  $q$  vanish at initial and terminal points, we get

$$\begin{aligned}
 & \int_0^T (|z'(t)|^2 + \frac{k}{2}|z(t)|^2) dt \\
 & \leq \int_0^T |z'(t)|^2 dt + \frac{1}{2} \int_0^T (|D_z H(t, (-A)^j z(t), (-A)^j \dot{q}(t))|^2 + |f_z(t)|^2) dt, \\
 (3.17) \quad &
 \end{aligned}$$

$$\begin{aligned}
 & \int_0^T (|q'(t)|^2 + \frac{k}{2}|q(t)|^2) dt \\
 & \leq \int_0^T |q'(t)|^2 dt + \frac{1}{2} \int_0^T (|D_q H(t, (-A)^j z(t), (-A)^j q(t))|^2 + |f_q(t)|^2) dt.
 \end{aligned}$$

Therefore this yields

$$\begin{aligned}
 & k \int_0^T (|z(t)|^2 + |q(t)|^2) dt \\
 (3.18) \quad & \leq \int_0^T (|f_z(t)|^2 + |f_q(t)|^2) dt + \int_0^T (|D_z H(t, (-A)^j z(t), (-A)^j q(t))|^2 \\
 & \quad + |D_q H(t, (-A)^j z(t), (-A)^j q(t))|^2) dt.
 \end{aligned}$$

Moreover, from the interpolation inequality (2.5) it follows

$$\begin{aligned}
 & |D_z H(t, (-A)^j z(t), (-A)^j q(t))| \\
 & = e^{\frac{k(t-T)^2}{2}} |D_p \bar{H}(t, (-A)^j \bar{y}(t), (-A)^j \bar{p}(t))| \\
 & \leq e^{\frac{k(t-T)^2}{2}} \theta(t) |D_p H(t, (-A)^j y_1(t), (-A)^j p_1(t))| \\
 (3.19) \quad & - D_p H(t, (-A)^j y_2(t), (-A)^j p_2(t))| \\
 & \leq L_H e^{\frac{k(t-T)^2}{2}} \theta(t) [|(-A)^j (y_1(t) - y_2(t))| + |(-A)^j (p_1(t) - p_2(t))|] \\
 & = L_H [|(-A)^j z(t)| + |(-A)^j q(t)|] \\
 & \leq L_H \sigma [|(-A)^{\frac{1}{2}} z(t)| + |(-A)^{\frac{1}{2}} q(t)|] + L_H C_\sigma [|z(t)| + |q(t)|]
 \end{aligned}$$



Similarly we have

$$(3.20) \quad |D_q H(t, (-A)^j z(t), (-A)^j q(t))| \\ \leq L_H \sigma[|(-A)^{\frac{1}{2}} z(t)| + |(-A)^{\frac{1}{2}} q(t)|] + L_H C'_\sigma[|z(t)| + |q(t)|].$$

Noticing that from estimates (3.19) and (3.20), for some positive constant  $C$  it follows

$$(3.21) \quad |D_z H(t, (-A)^j z(t), (-A)^j q(t))|^2 \\ \leq C L_H \sigma[|(-A)^{\frac{1}{2}} z(t)|^2 + |(-A)^{\frac{1}{2}} q(t)|^2] + C L_H C'_\sigma[|z(t)|^2 + |q(t)|^2],$$

and

$$(3.22) \quad |D_q H(t, (-A)^j z(t), (-A)^j q(t))|^2 \\ \leq C L_H \sigma[|(-A)^{\frac{1}{2}} z(t)|^2 + |(-A)^{\frac{1}{2}} q(t)|^2] + C L_H C'_\sigma[|z(t)|^2 + |q(t)|^2].$$

Multiplying by  $z(t)$  the first equation of (3.16) and integrating we obtain

$$\int_0^T |(-A)^{\frac{1}{2}} z(t)|^2 dt \\ = (kT + 1) \int_0^T |z(t)|^2 dt + \frac{1}{2} \int_0^T (|D_z H(t, (-A)^j z(t), (-A)^j q(t))|^2 + |f_z(t)|^2) dt.$$

Therefore from (3.21) it follows

$$(3.23) \quad \int_0^T |(-A)^{\frac{1}{2}} z(t)|^2 dt = (kT + 1) \int_0^T |z(t)|^2 dt \\ + \frac{C L_H \sigma}{2} \int_0^T (|(-A)^{\frac{1}{2}} z(t)|^2 + |(-A)^{\frac{1}{2}} q(t)|^2) dt + \frac{1}{2} \int_0^T |f_z(t)|^2 dt \\ + \frac{C L_H C'_\sigma}{2} \int_0^T (|z(t)|^2 + |q(t)|^2) dt.$$

Multiplying by  $-q(t)$  the second equation of (3.16) and integrating we obtain

$$\begin{aligned}
 (3.24) \quad & \int_0^T |(-A)^{\frac{1}{2}} q(t)|^2 dt \leq (kT + 1) \int_0^T |q(t)|^2 dt \\
 & + \frac{CL_H \sigma}{2} \int_0^T (|(-A)^{\frac{1}{2}} z(t)|^2 + |(-A)^{\frac{1}{2}} q(t)|^2) dt + \frac{1}{2} \int_0^T |f_q(t)|^2 dt \\
 & + \frac{CL_H C_\sigma}{2} \int_0^T (|z(t)|^2 + |q(t)|^2) dt.
 \end{aligned}$$

Adding (3.23) and (3.24), choosing  $0 < \sigma < \frac{1}{CL_H}$ , we have

$$\begin{aligned}
 (3.25) \quad & \int_0^T (|(-A)^{\frac{1}{2}} z(t)|^2 + |(-A)^{\frac{1}{2}} q(t)|^2) dt \\
 & \leq C_1 \int_0^T (|z(t)|^2 + |q(t)|^2) dt + \frac{1}{2(1 - CL_H \sigma)} \int_0^T (|f_z(t)|^2 + |f_q(t)|^2) dt,
 \end{aligned}$$

where

$$C_1 = \frac{kT + 1 + CL_H C_\sigma}{1 - CL_H \sigma}.$$

From estimates (3.18) and the above one, it follows

$$(3.26) \quad \int_0^T (|z(t)|^2 + |q(t)|^2) dt \leq \hat{C}(k) \int_0^T (|f_z(t)|^2 + |f_q(t)|^2) dt,$$

where  $\hat{C}(k)$  is positive for  $k$  big enough and  $0 < \sigma < \frac{1}{CL_H(1 + 2T)}$ . It is defined as

$$\hat{C}(k) = \frac{3 - 2CL_H \sigma}{2(1 - CL_H \sigma)(k + 2CL_H \sigma C_1 - 2CL_H C_\sigma)}.$$

notice that  $\hat{C} \rightarrow 0$  as  $k \rightarrow \infty$ .

From (3.26) directly follows

$$(3.27) \quad \int_0^T e^{k(t-T)^2} (|\bar{y}(t)|^2 + |\bar{p}(t)|^2) dt \leq \hat{C}(k) \int_0^T e^{k(t-T)^2} (|g_x(t)|^2 + |g_p(t)|^2) dt.$$

On the other hand

$$\begin{aligned}
 (3.28) \quad \int_0^T e^{k(t-T)^2} (|\bar{y}(t)|^2 + |\bar{p}(t)|^2) dt &\geq \int_0^{\frac{T}{2}} e^{k(t-T)^2} (|\hat{y}(t)|^2 + |\hat{p}(t)|^2) dt \\
 &\geq e^{k\frac{T^2}{4}} \int_0^{\frac{T}{2}} (|\hat{y}(t)|^2 + |\hat{p}(t)|^2) dt
 \end{aligned}$$

and the following holds

$$\begin{aligned}
 (3.29) \quad &\int_0^T e^{k(t-T)^2} (|g_x(t)|^2 + |g_p(t)|^2) dt \\
 &\leq \int_0^T e^{k(t-T)^2} |\theta'(t)|^2 (|\hat{y}(t)|^2 + |\hat{p}(t)|^2) dt \\
 &\leq \left(\frac{1}{\tilde{T}}\right)^2 \int_{\frac{T}{2}}^T e^{k(t-T)^2} (|\hat{y}(t)|^2 + |\hat{p}(t)|^2) dt \\
 &\leq \left(\frac{1}{\tilde{T}}\right)^2 e^{k\frac{T^2}{4}} \int_{\frac{T}{2}}^T (|\hat{y}(t)|^2 + |\hat{p}(t)|^2) dt.
 \end{aligned}$$

In conclusion, from (3.27), (3.28) and (3.29) we get

$$e^{k\frac{T^2}{4}} \int_0^{\frac{T}{2}} (|\hat{y}(t)|^2 + |\hat{p}(t)|^2) dt \leq c(k) \left(\frac{1}{\tilde{T}}\right)^2 e^{k\frac{T^2}{4}} \int_{\frac{T}{2}}^T (|\hat{y}(t)|^2 + |\hat{p}(t)|^2) dt.$$

From the above inequality we obtain

$$\int_0^{\frac{T}{2}} (|\hat{y}(t)|^2 + |\hat{p}(t)|^2) dt \rightarrow 0 \text{ as } k \rightarrow \infty.$$

and we conclude that  $|\hat{y}(t)| = |\hat{p}(t)| = 0$  on  $[0, \frac{T}{2}]$ . Iterating this procedure we obtain this result on  $[0, T]$ . ■

## References

1. V. Barbu, Boundary control problems with convex cost criterion, *SIAM J. Control and Optim.*, 18, 227-254, 1980.
2. V. Barbu and G. Da Prato, *Hamilton-Jacobi equations in Hilbert spaces*, Pitman, Boston, 1982.
3. A. Bensoussan, G. Da Prato, M. C. Delfour and S. K. Mitter, *Representation and control of infinite dimensional systems*, Birkhäuser, Boston, 1992.
4. P. Cannarsa, H. Frankowska, Value function and optimality conditions for semilinear control problems, *Applied Math. Optim.*, 26, 139-169, 1992.
5. P. Cannarsa and F. Gozzi, On the smoothness of the value function along optimal trajectories, *IFIP Workshop on Boundary Control and Boundary Variation* (to appear in *Lecture Notes in Control and Information Sciences*, Springer Verlag).
6. P. Cannarsa, F. Gozzi and H. M. Soner, A dynamic programming approach to nonlinear boundary control problems of parabolic type, *J. Funct. Anal.* (to appear).
7. P. Cannarsa and M. E. Tessitore, Cauchy problem for the dynamic programming equation of boundary control, *Proceedings IFIP Workshop on Boundary Control and Boundary Variation* (to appear).
8. F. Clarke and R. B. Vinter, The relationship between the maximum principle and dynamic programming, *SIAM J. Control Optim.*, 25, 1291-1311, 1989.
9. H. O. Fattorini, Boundary control system, *SIAM J. Control*, 6, 349-385, 1968.
10. H. O. Fattorini, A unified theory of necessary conditions for nonlinear nonconvex control systems, *Appl. Math. Opt.*, 15, 141-185, 1987.
11. A. Favini and A. Venni, On a two point problem for a system of abstract differential equations, *Numer. Funct. Anal. and Optimiz.*, 2(4), 301-322, 1980.
12. H. Goldberg and F. Tröltzsch, Second order sufficient optimality conditions for a class of nonlinear parabolic boundary control problems, *SIAM J. Control and Optim.*, vol. 31, 4, 1007-1025, 1993.
13. V. Ifode, Hamilton-Jacobi equations and boundary convex control problems, *Rev. Roumaine Math. Pures Appl.*, 34, (1989), 2, 117-127.
14. G. E. Ladas and V. Lakshmikantham, *Differential equations in abstract spaces*, Academic Press, New York and London, 1972.
15. I. Lasiecka and R. Triggiani, *Differential and algebraic equations with application to boundary control problems: continuous theory and approximation theory*, *Lecture Notes in Control and Information Sciences*, Springer-Verlag, 164, 1991.
16. J. L. Lions, *Optimal control of systems described by partial differential equations*, Springer-Verlag, Wiesbaden, 1972.
17. J. L. Lions and B. Malgrange, Sur l'unicité rétrograde dans les problèmes mixtes paraboliques, *Math. Scand.*, 8, 277-286, 1960.
18. A. Pazy, *Semigroups of linear operators and applications to partial differential equations*, Springer-Verlag, New York Heidelberg Berlin, 1983.
19. D. Preiss, Differentiability of Lipschitz functions on Banach spaces, *J. Funct. Anal.*, 91, 312-345, 1990.
20. F. Tröltzsch, On the semigroup approach for the optimal control of semilinear parabolic equations including distributed and boundary control, *Zeitschrift für Analysis und ihre Anwendungen Bd.*, 8, 431-443, 1989.

Piermarco Camarsa  
Dipartimento di Matematica  
Università di Roma "Tor Vergata"  
Via O. Raimondo  
I-00173 Roma, Italy

Maria Elisabetta Tessitore  
Dipartimento di Matematica  
Università di Roma "La Sapienza"  
Piazzale A. Moro 2  
I-00185 Roma, Italy

## PONTRYAGIN'S PRINCIPLE FOR OPTIMAL CONTROL PROBLEMS GOVERNED BY SEMILINEAR ELLIPTIC EQUATIONS

EDUARDO CASAS

Departamento de Matemática Aplicada y Ciencias de la Computación  
Universidad de Cantabria

**ABSTRACT.** In this paper we prove an extension of Pontryagin's principle for optimal control problems governed by semilinear elliptic partial differential equations. The control takes values in a bounded subset, not necessarily convex, of some Euclidean space; the cost functional is of Lagrange type; and some general equality and inequality state constraints are imposed. To derive Pontryagin's principle we combine a suitable penalization of the state constraints with Ekeland's principle. In the absence of equality state constraints we establish the optimality conditions in a qualified form for "almost all" problems. In a first stage, the classical spike perturbations of the controls used to derive Pontryagin's principle are replaced by a new type of variations.

1991 *Mathematics Subject Classification.* 49K20; Secondary 35J65

*Key words and phrases.* Pontryagin's principle, semilinear elliptic equations, state constraints, Ekeland's variational principle.

### 1. Introduction

Many papers dealing with optimality conditions for control problems governed by partial differential equations have been written in the last twenty five years, but only a few of them have been devoted to Pontryagin's principle. This makes a big difference with the theory of optimal control for ordinary differential equations, where this principle is the usual way of formulating the conditions for optimality. The motive of this difference is that the methods used for finite dimensional systems can not be readily extended to infinite systems, arising some difficulties in the way of this extension. Nevertheless some new ideas have appeared leading to the first results: let us mention the papers by Fattorini and Frankowska [14], Li and Yao [15] and Li and Yong [16] for problems governed by parabolic equations; Bonnans and Casas [3], [4], [5], Bonnans and Tiba [6], Casas and Yong [9] and Yong [18] for the elliptic case.

---

This research was partially supported by Dirección General de Investigación Científica y Técnica (Madrid).

This paper deals with Pontryagin's principle for state-constrained control problems governed by semilinear elliptic equations. The principle proved here is stronger than any of the previous ones corresponding to elliptic equations because of the generality of the state constraints or the assumptions on the data defining the problem. In particular, less regularity of the domain and the coefficients of the equation is required, and no continuity with respect to the control of the functions involved in the problem is assumed. Moreover some proofs have been simplified by using different methods. In spite of these remarks, this paper could be viewed as a combination of the ideas introduced in [5], [9] and [18].

Following [5], we will prove two minimum principles, called weak and strong respectively. The term strong is due to the fact of providing optimality conditions in a qualified form. This strong version of Pontryagin's principle is proved only for problems without equality state constraints and it requires an additional hypothesis: a stability condition of the optimal cost with respect to small perturbations of the feasible state set. This condition was first used by Bonnans [2]; see also [5] and [9]. Unlike [5], no stability assumption is needed here to prove the weak principle, this is achieved by using a different penalization of the state constraints, previously utilized in [9] and [18].

As mentioned above, the proof of the minimum principle is based on a penalization procedure of the state constraints. In this process, Ekeland's principle plays an essential role. Another important feature of the proof is the use of some special variations of the controls different of the classical spike perturbations. This idea was first employed by Li and Yuo [15] for evolution problems, and later it was adapted to elliptic problems in [18] and [9]. In this paper, we give a new proof of this adaptation, which is simpler and lets a better understanding of the nature of these perturbations.

The plan of this paper is as follows: in §2 we formulate the control problem and provide some examples; and in §3 and §4 we prove the weak and strong versions, respectively, of Pontryagin's principle.

## 2. Setting of the control problem

Let  $\Omega \subset \mathbb{R}^n$ ,  $n > 1$ , be an open bounded set, with boundary  $\Gamma$  of class  $C^1$  if  $n > 3$  and Lipschitz if  $n \leq 3$ . In  $\Omega$  we consider the differential operator

$$Ay = - \sum_{i,j=1}^n \partial_{x_i} [a_{ij}(x) \partial_{x_j} y],$$

where  $a_{ij} \in C(\bar{\Omega})$ ,  $1 \leq i, j \leq n$  and satisfy the usual ellipticity assumption: there exists  $\Lambda > 0$  such that

$$\sum_{i,j=1}^n a_{ij}(x) \xi_i \xi_j \geq \Lambda |\xi|^2 \quad \forall x \in \Omega \text{ and } \forall \xi \in \mathbb{R}^n.$$

Let  $K$  be a compact subset of  $\mathbb{R}^m$ ,  $m \geq 1$ , and  $f : \Omega \times \mathbb{R} \times K \rightarrow \mathbb{R}$  a Lebesgue measurable function of class  $C^1$  with respect to the second variable satisfying

$$(1) \quad \begin{cases} \frac{\partial f}{\partial y}(x, y, u) \leq 0 \quad \forall (x, y, u) \in \Omega \times \mathbb{R} \times K \\ \forall M > 0 \exists \psi_M \in L^s(\Omega), \quad s > n/2, \text{ such that} \\ |f(x, 0, u)| + \left| \frac{\partial f}{\partial y}(x, y, u) \right| \leq \psi_M(x) \quad \text{a.e. } x \in \Omega, \forall |y| \leq M, \forall u \in K. \end{cases}$$

If  $\Gamma$  is not of class  $C^1$ , but it is only Lipschitz and  $n = 2$ , then we also assume that  $s > 4/3$ .

Given a function  $u : \Omega \rightarrow K$ , we consider the state equation

$$(2) \quad \begin{cases} Ay = f(x, y(x), u(x)) & \text{in } \Omega \\ y = 0 & \text{on } \Gamma. \end{cases}$$

**Theorem 1.** *Under the previous assumptions, for every function  $u : \Omega \rightarrow K$ , such that  $x \in \Omega \rightarrow f(x, y, u(x)) \in \mathbb{R}$  is measurable for each  $y \in \mathbb{R}$ , there exists a number  $p > n$  such that (2) has a unique solution  $y_u$  in the Sobolev space  $W_0^{1,p}(\Omega)$ . Moreover there exists a constant  $C_K > 0$ , independent of  $u$ , such that*

$$(3) \quad \|y_u\|_{W_0^{1,p}(\Omega)} \leq C_K.$$

This theorem follows from the classical theory of partial differential equations. The first thing to remark is that  $f$  is a nonincreasing monotone function and it is enough to use a global or local-cutting procedure of  $f$  (see, for instance, [3] or [7]) to obtain the existence of a solution. The  $W_0^{1,p}(\Omega)$ -regularity follows from the fact that  $L^s(\Omega) \subset W^{-1,p}(\Omega)$  for every  $p \in (n, ns/(n-s)]$  if  $s < n$  and  $p \in (n, \infty)$  if  $s \geq n$ , and that operator  $A$  is an isomorphism from  $W_0^{1,p}(\Omega)$  onto  $W^{-1,p}(\Omega)$ ; see Morrey [17]. When  $n \leq 3$  and  $\Gamma$  is only Lipschitz, then the  $W_0^{1,p}(\Omega)$ -regularity is not always true, however there exists  $\epsilon > 0$  depending on  $\Omega$  such that for every

$$p \in \begin{cases} \left( \frac{4+\epsilon}{3+\epsilon}, 4+\epsilon \right) & \text{if } n = 2 \\ \left( \frac{3+\epsilon}{2+\epsilon}, 3+\epsilon \right) & \text{if } n = 3 \end{cases}$$

operator  $A : W_0^{1,p}(\Omega) \rightarrow W^{-1,p}(\Omega)$  is still an isomorphism; see Dahlberg [12].

An important consequence of the previous theorem is that  $y_u \in W_0^{1,p}(\Omega) \subset C_0(\Omega)$ , where  $C_0(\Omega)$  denotes the space of continuous functions in  $\bar{\Omega}$  vanishing on  $\Gamma$ . Moreover the previous inclusion is compact.

Established the fundamental result about the state equation, let us go to the formulation of the control problem. Firstly we consider a Lebesgue measurable



function  $L : \Omega \times \mathbb{R} \times K \longrightarrow \mathbb{R}$  of class  $C^1$  with respect to the second variable satisfying

$$(4) \quad \begin{cases} \forall M > 0 \quad \exists \phi_M \in L^1(\Omega) \quad \text{such that} \\ |L(x, 0, u)| + \left| \frac{\partial L}{\partial y}(x, y, u) \right| \leq \phi_M(x) \quad \text{a.e. } x \in \Omega, \forall |y| \leq M, \forall u \in K. \end{cases}$$

Let us denote by  $\mathcal{U}$  the set of admissible controls: measurable functions  $u : \Omega \rightarrow K$  such that

$$x \in \Omega \longrightarrow (f(x, y, u(x)), L(x, y, u(x))) \in \mathbb{R}^2 \quad \text{is measurable } \forall y \in \mathbb{R}.$$

Let  $Z$  be a separable Banach space and  $Q \subset Z$  a closed convex subset with nonempty interior. Finally, let  $G : W_0^{1,p}(\Omega) \longrightarrow Z$  and  $F : W_0^{1,p}(\Omega) \longrightarrow \mathbb{R}^l$ ,  $l \geq 1$ , be two mappings of class  $C^1$ . Then we can state the optimal control problem as follows

$$(P) \quad \begin{cases} \text{Minimize } J(u) = \int_{\Omega} L(x, y_u(x), u(x)) dx \\ u \in \mathcal{U}, G(y_u) \in Q, F(y_u) = 0. \end{cases}$$

Let us show how the usual examples of state constraints can be handled with this formulation.

**Example 1.** Given a continuous function  $g : \bar{\Omega} \times \mathbb{R} \longrightarrow \mathbb{R}$  of class  $C^1$  respect to the second variable and a number  $\delta > 0$ , with  $g(x, 0) < \delta \quad \forall x \in \Gamma$ , the constraint  $g(x, y_u(x)) \leq \delta$  for all  $x \in \Omega$  can be written in the above framework by putting  $Z = C_0(\Omega)$ ,  $G : W_0^{1,p}(\Omega) \longrightarrow C_0(\Omega)$ , defined by  $G(y) = g(\cdot, y(\cdot))$ , and

$$Q = \{z \in C_0(\Omega) : z(x) \leq \delta \quad \forall x \in \Omega\}.$$

**Example 2.** The constraint

$$\int_{\Omega} |y_u(x)| dx \leq \delta$$

is considered by taking  $Z = L^1(\Omega)$ ,  $G : W_0^{1,p}(\Omega) \longrightarrow L^1(\Omega)$ , with  $G(y) = y$ , and  $Q$  the closed ball in  $L^1(\Omega)$  of center at 0 and radius  $\delta$ .

**Example 3.** For every  $1 \leq j \leq k$  let  $g_j : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$  be a measurable function of class  $C^1$  with respect to the second variable such that for each  $M > 0$  there exists a function  $\eta_M^j \in L^1(\Omega)$  satisfying

$$|g_j(x, 0)| + \left| \frac{\partial g_j}{\partial y}(x, y) \right| \leq \eta_M^j(x) \quad \text{a.e. } x \in \Omega, \quad \forall |y| \leq M.$$

Then the constraints

$$\int_{\Omega} g_j(x, y_u(x)) dx = \delta_j, \quad 1 \leq j \leq k,$$

are included in the formulation of (P) by choosing  $G = (G_1, \dots, G_k)^T$ , with

$$G_j(y) = \int_{\Omega} g_j(x, y(x)) dx.$$

$Z = \mathbb{R}^k$ , and  $Q = (-\infty, \delta_1] \times \dots \times (-\infty, \delta_k]$ .

**Example 4.** The equality constraints

$$\int_{\Omega} f_j(x, y_u(x)) dx = 0, \quad 1 \leq j \leq l,$$

can also be included in problem (P) in the obvious way by assuming the same hypotheses as in Example 3.

**Example 5.** Let  $\{x_j\}_{j=1}^l \subset \Omega$ , then the equality constraints  $y_u(x_j) = \delta_j$ ,  $1 \leq j \leq l$ , are well adapted to the above formulation. Indeed, since  $W_0^{1,p}(\Omega) \subset C_0(\Omega)$ , the functions  $F_j : W_0^{1,p}(\Omega) \rightarrow \mathbb{R}$  given by  $F_j(y) = y(x_j) - \delta_j$  are well defined and they are of class  $C^1$ . Then it is enough to take  $F = (F_1, \dots, F_l)^T$ .

### 3. The weak Pontryagin's principle

The aim of this section is to prove the following result

**Theorem 2.** Let  $p > n$  be given by Theorem 1. If  $\bar{u}$  is a solution of (P), then there exist  $\bar{\alpha} \geq 0$ ,  $\bar{y} \in W_0^{1,p'}(\Omega)$ ,  $\bar{\varphi} \in W_0^{1,p'}(\Omega)$ ,  $p' = p/(p-1)$ ,  $\bar{\mu} \in Z'$  and  $\bar{\lambda} \in \mathbb{R}^l$  such that

$$(5) \quad \bar{\alpha} + \|\bar{\mu}\|_{Z'} + |\bar{\lambda}| > 0;$$

$$(6) \quad \begin{cases} A\bar{y} = f(x, \bar{y}(x), \bar{u}(x)) & \text{in } \Omega, \\ \bar{y} = 0 & \text{on } \Gamma; \end{cases}$$

$$(7) \quad \begin{cases} A^* \bar{\varphi} = \frac{\partial f}{\partial y}(x, \bar{y}(x), \bar{u}(x)) \bar{\varphi} + \bar{\alpha} \frac{\partial L}{\partial y}(x, \bar{y}(x), \bar{u}(x)) \\ \quad + [DG(\bar{y})]^* \bar{\mu} + [DF(\bar{y})]^* \bar{\lambda} & \text{in } \Omega, \\ \bar{\varphi} = 0 & \text{on } \Gamma; \end{cases}$$

$$(8) \quad \langle \bar{\mu}, z - G(\bar{y}) \rangle \leq 0 \quad \forall z \in Q;$$

$$(9) \quad \int_{\Omega} H_{\bar{\alpha}}(x, \bar{y}(x), \bar{u}(x), \bar{\varphi}(x)) dx = \min_{v \in \mathcal{U}} \int_{\Omega} H_{\bar{\alpha}}(x, \bar{y}(x), v(x), \bar{\varphi}(x)) dx;$$

where

$$H_\alpha(x, y, u, \varphi) = \alpha L(x, y, u) + \varphi f(x, y, u).$$

Moreover, if one of the following assumptions is satisfied:

(A1) Functions  $f$  and  $L$  are continuous with respect to the third variable on  $K$ ;

(A2) There exists a set  $\Omega_0 \subset \Omega$ , with  $|\Omega_0| = |\Omega|$ , such that the function

$$x \in \Omega \rightarrow (f(x, y, u), L(x, y, u)) \in \mathbb{R}^2$$

is continuous in  $\Omega_0$  for every  $(y, u) \in \mathbb{R} \times K$ ;

then the following pointwise relation holds

$$(10) \quad H_\alpha(x, y(x), u(x), \varphi(x)) = \min_{v \in K} H_\alpha(x, y(x), v, \varphi(x)) \quad \text{a.e. } x \in \Omega.$$

Let us see how the previous optimality conditions are written for some of the state constraints given in the examples of §2.

**Example 6.** Let us assume that the constraints associated with  $G$  and  $F$  are those given in the examples 1 and 4, respectively. Then, taking into account that  $Z' = C_0(\Omega)'$  is identified with the space of real and regular Borel measures in  $\Omega$ , denoted by  $M(\Omega)$ , Theorem 2 claims the existence of a measure  $\mu \in M(\Omega)$ , a number  $\alpha \geq 0$ , a vector  $\lambda \in \mathbb{R}^l$  and two functions  $y \in W_0^{1,p}(\Omega)$  and  $\varphi \in W_0^{1,p'}(\Omega)$  such that (5), (6), (9) and possibly (10), and the following relations hold

$$\begin{cases} A^* \varphi = \frac{\partial f}{\partial y}(x, y(x), u(x)) \varphi + \alpha \frac{\partial L}{\partial y}(x, y(x), u(x)) \\ \quad + \frac{\partial g}{\partial y}(x, y(x)) \mu + \sum_{j=1}^l \lambda_j \frac{\partial f_j}{\partial y}(x, y(x)) \quad \text{in } \Omega, \\ \varphi = 0 \quad \text{on } \Gamma; \end{cases}$$

$$\int_{\Omega} (z(x) - g(x, y(x))) d\mu(x) \leq 0 \quad \forall z \in C_0(\Omega), \quad \text{with } z(x) \leq \delta \quad \forall x \in \Omega.$$

**Example 7.** If the state constraints are given as in the examples 2 and 5, then  $\bar{\mu} \in L^\infty(\Omega)$  and the adjoint state equation (7) and inequality (8) become now

$$\begin{cases} A^* \bar{\varphi} = \frac{\partial f}{\partial y}(x, \bar{y}(x), \bar{u}(x)) \bar{\varphi} + \alpha \frac{\partial L}{\partial y}(x, \bar{y}(x), \bar{u}(x)) + \bar{\mu} + \sum_{j=1}^l \lambda_j \delta_{[x_j]} \quad \text{in } \Omega, \\ \bar{\varphi} = 0 \quad \text{on } \Gamma; \end{cases}$$

$$\int_{\Omega} \bar{\mu}(x)(z(x) - \bar{y}(x)) dx \leq 0 \quad \forall z \in L^1(\Omega) \quad \text{with} \quad \int_{\Omega} |z(x)| dx \leq \delta;$$

where  $\delta_{[x_j]}$  denotes the Dirac measure centered at  $x_j$ .

Finally, let us remark that if the constraints associated to  $G$  are given as in Example 3, then  $\mu = \{\mu_j\}_{j=1}^k \in \mathbb{R}^k$  and (8) states that  $\mu_j \geq 0$  for every  $j = 1, \dots, k$ .

Now we go on to prove Theorem 2. First of all we need two lemmas, which describe the type of variations of the controls used in the proof.

**Lemma 1.** *Let  $g \in L^1(\Omega)$  and  $h \in L^s(\Omega)$ ,  $1 \leq s \leq +\infty$ . Given  $\rho \in (0, 1)$ , let us denote*

$$\mathcal{E}_\rho = \{E \subset \Omega \text{ measurable} : |E| = \rho|\Omega|\}.$$

*Then*

$$\inf_{E \in \mathcal{E}_\rho} \left\{ \left| \int_\Omega \left(1 - \frac{1}{\rho} \chi_E(x)\right) g(x) dx \right| + \left\| \left(1 - \frac{1}{\rho} \chi_E\right) h \right\|_{W^{1,p}(\Omega)} \right\} = 0,$$

*where  $\chi_E$  is the characteristic function of  $E$  and  $p$  is any real number satisfying*

$$\begin{cases} 1 < p \leq +\infty & \text{if } s \geq n, \\ 1 < p \leq ns/(n-s) & \text{if } s < n. \end{cases}$$

*Proof.* Given  $\rho \in (0, 1)$ , let us take  $\epsilon > 0$  arbitrary. Let us denote by  $B$  the closed unit ball of  $W_0^{1,p'}(\Omega)$ , where  $p' = p/(p-1)$ . From the assumptions on  $p$  and the Sobolev's imbeddings (see, for instance, Adams [1]) it follows

$$L^s(\Omega) \subset W^{-1,p}(\Omega) \quad \text{and} \quad W_0^{1,p'}(\Omega) \subset L^{s'}(\Omega),$$

the imbeddings being continuous. Above we denote  $s' = s/(s-1)$ . Let us assume that  $s' < +\infty$ , otherwise it is enough to change the integrals by the supremum norm to accomplish the proof.

We can take a partition  $\{\Omega_j\}_{j=1}^r$  of  $\Omega$ , with  $|\Omega_j| > 0$ , such that

$$(11) \quad \sum_{j=1}^r \int_{\Omega_j} \left| y(x) - \frac{1}{|\Omega_j|} \int_{\Omega_j} y(\xi) d\xi \right|^{s'} dx < \epsilon^{s'} \quad \forall y \in B.$$

This can be proved, for instance, by using the interpolation theory in Sobolev spaces; see Ciarlet [10].

For every  $y \in B$  let us define  $\tilde{y} : \Omega \rightarrow \mathbb{R}$  by

$$\tilde{y}|_{\Omega_j}(x) = \frac{1}{|\Omega_j|} \int_{\Omega_j} y(\xi) d\xi \quad \forall x \in \Omega_j, \quad 1 \leq j \leq r.$$

Then (11) implies

$$(12) \quad \|y - \tilde{y}\|_{L^{s'}(\Omega)} < \epsilon \quad \forall y \in B.$$

Now we can approximate  $g$  and  $h$  in every  $\Omega_j$  by simple functions

$$g_j(x) = \sum_{i=1}^{r_j} \alpha_{ij} \chi_{F^{ij}}(x), \quad h_j(x) = \sum_{i=1}^{r_j} \beta_{ij} \chi_{F^{ij}}(x),$$

with  $\alpha_{ij}, \beta_{ij} \in \mathbb{R}$ ,  $\{F^{ij}\}_{i=1}^{r_j}$  being a partition of  $\Omega_j$  with  $|F^{ij}| > 0$ , and such that

$$(13) \quad \int_{\Omega_j} |g(x) - g_j(x)| dx + \int_{\Omega_j} |h(x) - h_j(x)| dx < \epsilon |\Omega_j|.$$

Let us take  $E^{ij} \subset F^{ij}$  such that  $|E^{ij}| = \rho |F^{ij}|$ . Defining  $E^j = \bigcup_{i=1}^{r_j} E^{ij}$ , we have that

$$(14) \quad \int_{\Omega_j} g_j(x) dx = \int_{\Omega_j} \frac{1}{\rho} \chi_{E^j}(x) g_j(x) dx, \quad 1 \leq j \leq r,$$

and

$$(15) \quad \int_{\Omega_j} h_j(x) dx = \int_{\Omega_j} \frac{1}{\rho} \chi_{E^j}(x) h_j(x) dx, \quad 1 \leq j \leq r.$$

Finally we take  $E = \bigcup_{j=1}^r E^j$ . Then  $|E| = \rho |\Omega|$  and for every  $y \in B$  we have

$$(16) \quad \begin{aligned} \left| \int_{\Omega} \left( 1 - \frac{1}{\rho} \chi_E(x) \right) h(x) y(x) dx \right| &\leq \left| \int_{\Omega} \left( 1 - \frac{1}{\rho} \chi_E(x) \right) h(x) \tilde{y}(x) dx \right| \\ &\quad + \int_{\Omega} \left( 1 + \frac{1}{\rho} \chi_E(x) \right) |h(x)| |y(x) - \tilde{y}(x)| dx \\ &\leq \left| \sum_{j=1}^r \int_{\Omega_j} \left( 1 - \frac{1}{\rho} \chi_{E^j}(x) \right) h(x) \tilde{y}(x) dx \right| \\ &\quad + \left( 1 + \frac{1}{\rho} \right) \|h\|_{L^s(\Omega)} \|y - \tilde{y}\|_{L^{s'}(\Omega)}. \end{aligned}$$

From (11) we deduce

$$(17) \quad \left( 1 + \frac{1}{\rho} \right) \|h\|_{L^s(\Omega)} \|y - \tilde{y}\|_{L^{s'}(\Omega)} \leq \epsilon \left( 1 + \frac{1}{\rho} \right) \|h\|_{L^s(\Omega)}.$$

On the other hand, from (13) and (15) we obtain for all  $y \in B$

$$\begin{aligned}
 \left| \sum_{j=1}^r \int_{\Omega_j} \left(1 - \frac{1}{\rho} \chi_E(x)\right) h(x) \tilde{y}(x) dx \right| &\leq \left(1 + \frac{1}{\rho}\right) \sum_{j=1}^r \left\{ |\tilde{y}|_{\Omega_j} \int_{\Omega_j} |h(x) - h_j(x)| dx \right\} \\
 (18) \quad &+ \left| \sum_{j=1}^r |\tilde{y}|_{\Omega_j} \int_{\Omega_j} \left(1 - \frac{1}{\rho} \chi_E(x)\right) h_j(x) dx \right| \\
 &\leq \left(1 + \frac{1}{\rho}\right) \epsilon \|\tilde{y}\|_{L^1(\Omega)} \leq C \left(1 + \frac{1}{\rho}\right) \epsilon.
 \end{aligned}$$

Thus (16)–(18) imply that

$$(19) \quad \left\| \left(1 - \frac{1}{\rho} \chi_E\right) h \right\|_{W^{1,p}(\Omega)} \leq \epsilon \left(1 + \frac{1}{\rho}\right) (C + \|h\|_{L^1(\Omega)}),$$

where we have used that  $W^{1,p}(\Omega) = \left(W_0^{1,p}(\Omega)\right)'$ .

On the other hand, using (13) and (14) we get

$$\begin{aligned}
 \left| \int_{\Omega} \left(1 - \frac{1}{\rho} \chi_E\right) g(x) dx \right| &\leq \sum_{j=1}^r \left| \int_{\Omega_j} \left(1 - \frac{1}{\rho} \chi_E(x)\right) g(x) dx \right| \\
 (20) \quad &\leq \left(1 + \frac{1}{\rho}\right) \sum_{j=1}^r \int_{\Omega_j} |g(x) - g_j(x)| dx \\
 &+ \sum_{j=1}^r \left| \int_{\Omega_j} \left(1 - \frac{1}{\rho} \chi_E(x)\right) g_j(x) dx \right| \\
 &\leq \epsilon \left(1 + \frac{1}{\rho}\right) |\Omega|.
 \end{aligned}$$

Then (19) and (20), together with the fact that  $\epsilon > 0$  is arbitrary, prove the lemma.  $\square$

**Lemma 2.** Suppose  $p > n$  given by Theorem 1. Let  $u \in \mathcal{U}$  and let  $y$  be its associated state. Then for every  $\rho \in (0, 1)$  we can find a measurable set  $E_\rho \subset \Omega$ , with  $|E_\rho| = \rho|\Omega|$ , such that if we define

$$u_\rho(x) = \begin{cases} u(x) & \text{if } x \in \Omega \setminus E_\rho \\ v(x) & \text{if } x \in E_\rho, \end{cases}$$

with  $v \in \mathcal{U}$ , and if we denote by  $y_\rho$  the state corresponding to  $u_\rho$ , the following equalities hold

$$(21) \quad y_\rho = y + \rho z + r_\rho, \quad \lim_{\rho \rightarrow 0} \frac{1}{\rho} \|r_\rho\|_{W_0^{1,p}(\Omega)} = 0,$$

and

$$(22) \quad J(u_\rho) = J(u) + \rho z^0 + r_\rho^0, \quad \lim_{\rho \rightarrow 0} \frac{1}{\rho} |r_\rho^0| = 0,$$

where  $z \in W_0^{1,p}(\Omega)$  satisfies

$$(23) \quad Az = \frac{\partial f}{\partial y}(x, y, u)z + f(x, y, v) - f(x, y, u) \quad \text{in } \Omega$$

and

$$(24) \quad z^0 = \int_{\Omega} \left\{ \frac{\partial L}{\partial y}(x, y(x), u(x))z(x) + L(x, y(x), v(x)) - L(x, y(x), u(x)) \right\} dx.$$

*Proof.* Let us define

$$g(x) = L(x, y(x), v(x)) - L(x, y(x), u(x))$$

and

$$h(x) = f(x, y(x), v(x)) - f(x, y(x), u(x)).$$

Then, thanks to the hypotheses (1) and (4) we have that  $h \in L^s(\Omega)$  and  $g \in L^1(\Omega)$ . Then we can apply Lemma 1 to obtain the existence of a measurable set  $E_\rho \subset \Omega$ , with  $|E_\rho| = \rho|\Omega|$ , satisfying

$$(25) \quad \left| \int_{\Omega} \left( 1 - \frac{1}{\rho} \chi_{E_\rho}(x) \right) g(x) dx \right| + \left\| \left( 1 - \frac{1}{\rho} \chi_{E_\rho} \right) h \right\|_{W^{-1,p}(\Omega)} \leq \rho.$$

Putting

$$z_\rho = \frac{y_\rho - y}{\rho},$$

it is easy to obtain, by subtracting the equations satisfied by  $y_\rho$  and  $y$ , respectively, that

$$(26) \quad Az_\rho = a^\rho(x)z_\rho + \frac{1}{\rho} \chi_{E_\rho} h \quad \text{in } \Omega,$$

where

$$a^\rho(x) = \int_0^1 \frac{\partial f}{\partial y}(x, y(x) + \theta[y_\rho(x) - y(x)], u_\rho(x)) d\theta.$$

It is not difficult to check with the help of the  $W_0^{1,p}(\Omega)$  regularity of  $A$ , that

$$(27) \quad \|y_\rho - y\|_{W_0^{1,p}(\Omega)} \longrightarrow 0.$$

In particular, because of the imbedding  $W_0^{1,p}(\Omega) \subset C(\bar{\Omega})$ , we deduce from (27) together with (1) that

$$(28) \quad a^\rho(x) \longrightarrow \frac{\partial f}{\partial y}(x, y(x), u(x)) \quad \text{in } L^s(\Omega).$$

This relation, (26) and the convergence  $(1/\rho)\chi_{E_\rho} h \rightarrow h$  in  $W^{1,p}(\Omega)$  implies that  $z_\rho \rightarrow z$  in  $W_0^{1,p}(\Omega)$ , which proves (21).

On the other hand

$$\begin{aligned} \frac{1}{\rho}|r_\rho^0| &= \left| \frac{J(u_\rho) - J(u)}{\rho} - z^0 \right| \\ &\leq \int_\Omega \left| \int_0^1 \frac{\partial L}{\partial y}(x, y(x) + \theta[y_\rho(x) - y(x)], u_\rho(x)) d\theta z_\rho(x) - \frac{\partial L}{\partial y}(x, y(x), u(x)) z(x) \right| dx \\ &\quad + \left| \int_\Omega \left( 1 - \frac{1}{\rho} \chi_{E_\rho}(x) \right) g(x) dx \right| \longrightarrow 0 \quad \text{when } \rho \rightarrow 0, \end{aligned}$$

which proves (22).  $\square$

*Proof of Theorem 2.*

Since  $Z$  is separable, we can take in  $Z$  a norm  $\|\cdot\|_Z$  such that  $Z'$  endowed with the dual norm  $\|\cdot\|_{Z'}$  is strictly convex. Then the function

$$d_Q : (Z, \|\cdot\|_Z) \longrightarrow \mathbb{R}$$

$$d_Q(z) = \inf_{y \in Q} \|y - z\|_Z$$

is convex, Lipschitz and Gâteaux differentiable at every point  $z \notin Q$ , with  $\partial d_Q(z) = \{\nabla d_Q(z)\}$ , where the Clarke's generalized gradient and the subdifferential in the sense of the convex analysis coincides for this function. Therefore, given  $\xi \in \partial d_Q(y)$ , we have that

$$(29) \quad \langle \xi, z - y \rangle + d_Q(y) \leq d_Q(z) \quad \forall z \in Z.$$

Moreover  $\|\nabla d_Q(z)\|_{Z'} = 1$  for every  $z \notin Q$ ; see Clarke [11] and Casas and Yong [9].

Let us take  $J_\epsilon : \mathcal{U} \longrightarrow \mathbb{R}$  defined by

$$J_\epsilon(u) = \{[(J(u) - J(u) + \epsilon)^+]^2 + d_Q(G(y_u))^2 + |F(y_u)|^2\}^{1/2}.$$

It is obvious that  $J_\epsilon(u) > 0$  for every  $u \in \mathcal{U}$  and  $J_\epsilon(u) = \epsilon$ . On the other hand, if  $d$  denotes the Ekeland's distance in  $\mathcal{U}$ :

$$d(u, v) = |\{x \in \Omega : u(x) \neq v(x)\}|,$$

then  $J_\epsilon$  is continuous in  $(\mathcal{U}, d)$ . Therefore we can apply Ekeland's variational principle [13] and deduce the existence of  $u' \in \mathcal{U}$  such that

$$(30) \quad d(u', \bar{u}) \leq \sqrt{\epsilon} \quad \text{and} \quad 0 < J_\epsilon(u') \leq J_\epsilon(u) + \sqrt{\epsilon} d(u', u) \quad \forall u \in \mathcal{U}.$$

Taking  $E_\rho$  and  $u'_\rho$  as in Lemma 2,

$$u'_\rho(x) = \begin{cases} u'(x) & \text{if } x \in \Omega \setminus E_\rho \\ v(x) & \text{if } x \in E_\rho, \end{cases}$$



with  $v \in \mathcal{U}$  arbitrary, we get with the help of (21) and (22)

$$\begin{aligned}
 -\sqrt{\epsilon}|\Omega| &\leq \frac{J_\epsilon(u'_\rho) - J_\epsilon(u')}{\rho} \\
 &= \frac{[(J(u'_\rho) - J(u) + \epsilon)^4]^2 - [(J(u') - J(u) + \epsilon)^4]^2}{\rho[J_\epsilon(u'_\rho) + J_\epsilon(u')]} \\
 (31) \quad &+ \frac{d_Q(G(y'_\rho))^2 - d_Q(G(y'))^2 + |F(y'_\rho)|^2 - |F(y')|^2}{\rho[J_\epsilon(u'_\rho) + J_\epsilon(u')]} \\
 &\xrightarrow{\rho \rightarrow 0} \frac{\{(J(u') - J(u) + \epsilon)^4 z^{0,\epsilon} + \langle \xi', DG(y')z' \rangle + \langle F(y'), DF(y')z' \rangle\}}{J_\epsilon(u')} \\
 &= \alpha_\epsilon z^{0,\epsilon} + \langle [DG(y')]^* \mu', z' \rangle + \langle [DF(y')]^* \lambda', z' \rangle,
 \end{aligned}$$

where  $z' \in W_0^{1,p}(\Omega)$ ,

$$\begin{aligned}
 Az' &= \frac{\partial f}{\partial y}(x, y', u')z' + f(x, y', v) - f(x, y', u'), \\
 z^{0,\epsilon} &= \int_\Omega \left\{ \frac{\partial L}{\partial y}(x, y', u')z' + L(x, y', v) - L(x, y', u') \right\} dx, \\
 \alpha_\epsilon &= \frac{(J(u') - J(u) + \epsilon)^4}{J_\epsilon(u')}, \quad \mu' = \frac{\xi_\epsilon}{J_\epsilon(u')}, \quad \lambda' = \frac{F(y')}{J_\epsilon(u')}, \\
 \xi' &= \begin{cases} d_Q(G(y')) \nabla d_Q G(y') & \text{if } G(y') \notin Q \\ 0 & \text{otherwise.} \end{cases}
 \end{aligned}$$

Now let us remind that  $A : W_0^{1,p}(\Omega) \rightarrow W^{1,p}(\Omega)$  is an isomorphism; see the comments after Theorem 1. Then  $A^* : W_0^{1,p'}(\Omega) \rightarrow W^{1,p'}(\Omega)$  is also an isomorphism, therefore we can introduce the adjoint state  $\varphi' \in W_0^{1,p'}(\Omega)$  as solution of the equation

$$(32) \quad A^* \varphi' = \frac{\partial f}{\partial y}(x, y', u') \varphi' + \alpha_\epsilon \frac{\partial L}{\partial y}(x, y', u') + [DG(y')]^* \mu' + [DF(y')]^* \lambda'.$$

Thus, from (31) and (32), we obtain

$$\begin{aligned}
 (33) \quad \int_\Omega [H_{\alpha_\epsilon}(x, y'(x), u'(x), \varphi'(x)) - H_{\alpha_\epsilon}(x, y'(x), v(x), \varphi'(x))] dx \\
 \leq \sqrt{\epsilon}|\Omega| \quad \forall v \in \mathcal{U}.
 \end{aligned}$$

Now we pass to the limit when  $\epsilon \rightarrow 0$ . To do this, let us remark that

$$(34) \quad \alpha_\epsilon^2 + \|\mu'\|_{Z'}^2 + |\lambda'|^2 = 1.$$

Then we take subsequences, denoted in the same way, satisfying

$$(35) \quad \begin{cases} \alpha_\epsilon \rightarrow \bar{\alpha} \text{ in } \mathbb{R}, & \lambda' \rightarrow \lambda \text{ in } \mathbb{R}^n \\ \mu' \rightarrow \bar{\mu} \text{ in the weak}^* \text{ topology of } Z'. \end{cases}$$

On the other hand, the convergence  $y' \rightarrow y$  in  $W_0^{1,p}(\Omega) \subset C_0(\Omega)$  follows from (30). Then, using (35), it is easy to pass to the limit in (32) and (33) and to deduce (7) and (9). Now reminding the definition of  $\mu'$  and  $\xi'$  and (29), we deduce

$$(36) \quad \langle \mu', z - G(y') \rangle \leq 0 \quad \forall z \in Q.$$

Passing to the limit in this expression we obtain (8). Let us prove (5). To do this, let us suppose that  $\bar{\alpha} = |\bar{\lambda}| = 0$ , then from (34) it follows  $\|\mu'\|_{Z'} \rightarrow 1$  as  $\epsilon \rightarrow 0$ .

Let us take  $z_0 \in \overset{\circ}{Q}$  and  $r > 0$  such that  $\bar{B}_r(z_0) \subset \overset{\circ}{Q}$ . Then (36) implies that

$$\langle \mu', z + z_0 - G(y') \rangle \leq 0 \quad \forall z \in \bar{B}_r(z_0).$$

Hence

$$r\|\mu'\|_{Z'} = \sup_{z \in B_r(z_0)} \langle \mu', z \rangle \leq \langle \mu', G(y') - z_0 \rangle.$$

Passing to the limit

$$0 < r \leq \lim_{\epsilon \rightarrow 0} \langle \mu', G(y') - z_0 \rangle = \langle \mu, G(\bar{y}) - z_0 \rangle,$$

which proves that  $\mu \neq 0$ .

Finally let us deduce (10) from (9). Assume firstly that (A1) holds. Let us take a numerable dense subset  $\{v_j\}_{j=1}^\infty$  of  $K$ . Let  $F$  and  $\{F_j\}_{j=1}^\infty$  be measurable subsets of  $\Omega$ , with  $|F| = |\Omega| = |F_j|$  for every  $j$ , such that the Lebesgue point sets of functions  $x \in \Omega \rightarrow H(x, \bar{y}(x), u(x), \varphi(x))$  and  $x \in \Omega \rightarrow H(x, \bar{y}(x), v_j, \varphi(x))$

are  $F$  and  $F_j$ , respectively. Let us set  $F_0 = F \bigcap \bigcap_{j=1}^\infty F_j$ . Then we have  $|F_0| = |\Omega|$ .

Now given  $x_0 \in F_0$  arbitrary, for every  $\epsilon > 0$  and  $j \geq 1$  we define the admissible controls

$$u'_j(x) = \begin{cases} u(x) & \text{if } x \notin B_\epsilon(x_0) \\ v_j & \text{otherwise.} \end{cases}$$

Then from (9) we deduce

$$\begin{aligned} \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} H_\alpha(x, y(x), u(x), \varphi(x)) dx \\ \leq \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} H_\alpha(x, \bar{y}(x), v_j, \varphi(x)) dx, \quad 1 \leq j. \end{aligned}$$

Passing to the limit when  $\epsilon \rightarrow 0$  we get

$$H_\alpha(x_0, \bar{y}(x_0), \bar{u}(x_0), \bar{\varphi}(x_0)) \leq H_\alpha(x_0, \bar{y}(x_0), v_j, \bar{\varphi}(x_0)) \quad \forall x_0 \in F_0, \quad \forall j \geq 1.$$

Taking into account that  $v \rightarrow H_\alpha(x_0, \bar{y}(x_0), v, \bar{\varphi}(x_0))$  is continuous and that  $\{v_j\}_{j=1}^\infty$  is dense in  $K$ , (10) follows from the above inequality.

Now let us suppose that assumption (A2) holds. Let  $F_\varphi$  be a measurable subset of  $\Omega$  such that

$$(37) \quad \lim_{\epsilon \rightarrow 0} \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} |\varphi(x) - \varphi(x_0)|^{s'} dx = 0 \quad \forall x_0 \in F_\varphi,$$

where  $s' = s/(s-1)$ ,  $s$  given in (1). Then  $|F_\varphi| = |\Omega|$ . Let  $M = \|\bar{y}\|_\infty$  and  $\psi_M \in L^s(\Omega)$  satisfying (1). Finally we put  $F_0 = F_\varphi \cap F_M \cap \Omega_0 \cap F$ , where  $F$  is taken as above and  $F_M$  is the set of Lebesgue points of  $|\psi_M|^s$ . Thus we have that  $|F_0| = |\Omega|$  and taking spike perturbations as before, we deduce

$$\begin{aligned} & \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} H_\alpha(x, \bar{y}(x), \bar{u}(x), \varphi(x)) dx \\ & \leq \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} H_\alpha(x, \bar{y}(x), v, \bar{\varphi}(x)) dx \quad \forall x_0 \in F_0 \text{ and } \forall v \in K. \end{aligned}$$

Since  $x_0 \in F$ , we can pass to the limit on the left hand side of the inequality. Let us study the right hand side.

$$\begin{aligned} & \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} H_\alpha(x, \bar{y}(x), v, \bar{\varphi}(x)) dx \\ & = \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} \bar{\alpha} L(x, \bar{y}(x), v) dx + \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} f(x, \bar{y}(x), v) dx \varphi(x_0) \\ & \quad + \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} [\varphi(x) - \varphi(x_0)] f(x, \bar{y}(x), v) dx. \end{aligned}$$

The first two terms converge to  $H_\alpha(x_0, \bar{y}(x_0), v, \bar{\varphi}(x_0))$  because of the continuity of the integrands in  $x_0$ . Let us prove that the last term goes to zero.

$$\begin{aligned} & \left| \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} [\varphi(x) - \varphi(x_0)] f(x, \bar{y}(x), v) dx \right| \\ & \leq \left( \frac{1}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} |\varphi(x) - \varphi(x_0)|^{s'} dx \right)^{1/s'} \left( \frac{1+M}{|B_\epsilon(x_0)|} \int_{B_\epsilon(x_0)} |\psi_M(x)|^s dx \right)^{1/s} \rightarrow 0, \end{aligned}$$

thanks to (37) and the fact that  $x_0 \in F_\varphi \cap F_M$ .  $\square$

#### 4. The strong Pontryagin's principle

In this section we will prove that, in the absence of equality constraints, Theorem 2 holds with  $\bar{\alpha} = 1$  for "almost all" control problems. We will precise this term later. The key to achieve this result is the introduction of a stability assumption of the optimal cost functional with respect to small perturbations of the set of feasible controls. This stability allows to accomplish an exact penalization

of the state constraints. First of all let us formulate the control problem to be studied in this section

$$(P_\delta) \begin{cases} \text{Minimize } J(u) \\ u \in \mathcal{U}, G(y_u) \in Q_\delta \end{cases}$$

with the same notation and assumptions of §2 and setting  $Q_\delta = Q + \bar{B}_\delta(0)$ , for every  $\delta > 0$ .

**Definition 1.** We say that  $(P_\delta)$  is strongly stable if there exist  $\epsilon > 0$  and  $C' > 0$  such that

$$(38) \quad \inf(P_\delta) - \inf(P_{\delta'}) \leq C'(\delta' - \delta) \quad \forall \delta' \in [\delta, \delta + \epsilon].$$

This concept was first introduced in relation with optimal control problems by Bonnans [2]; see also Bonnans and Casas [5]. A weaker stability concept was used by Casas [8] to analyze the convergence of the numerical discretizations of optimal control problems. The following proposition states that almost all problems  $(P_\delta)$  are strongly stable.

**Proposition 1.** Let  $\delta_0 \geq 0$  be a number such that  $(P_{\delta_0})$  has feasible controls. Then  $(P_\delta)$  is strongly stable for all  $\delta \geq \delta_0$  except at most a zero Lebesgue measure set.

*Proof.* It is enough to consider the function  $h : [\delta_0, +\infty) \rightarrow \mathbb{R}$  defined by

$$h(\delta) = \inf(P_\delta)$$

and remark that it is a nonincreasing monotone function and, consequently, differentiable at every point of  $[\delta_0, +\infty)$  except at a zero measure set. Now it is obvious to check that  $(P_{\delta_0})$  is strongly stable at every point where  $h$  is differentiable.  $\square$

Now we carry out an exact penalization of the state constraint. To do this, we will use the distance function  $d_{Q_\delta}$  associated to the set  $Q_\delta$ , and define in the same way as in the proof of Theorem 2.

**Proposition 2.** If  $(P_\delta)$  is strongly stable and  $\bar{u}$  is a solution of this problem, then there exists  $q_0 > 0$  such that  $\bar{u}$  is also a solution of

$$(39) \quad \inf_{u \in \mathcal{U}} J_q(u) = J(\bar{u}) + q d_{Q_\delta}(G(y_{\bar{u}}))$$

for every  $q \geq q_0$ .

*Proof.* Let us suppose that it is false. Then there exists a sequence  $\{q_k\}_{k=1}^\infty$  of real numbers, with  $q_k \rightarrow +\infty$ , and elements  $\{u_k\}_{k=1}^\infty \subset \mathcal{U}$  such that

$$J(u_k) + q_k d_{Q_\delta}(G(y_k)) < J(\bar{u}) \quad \forall k \geq 1,$$

where  $y_k$  is the state corresponding to  $u_k$ . From here we obtain that

$$d_{Q_\delta}(G(y_k)) < \frac{J(\bar{u}) - J(u_k)}{q_k} \rightarrow 0 \quad \text{when } k \rightarrow +\infty$$

and  $G(y_k) \notin Q_\delta$ . Let  $\delta_k > \delta$  be the smallest number such that  $G(y_k) \in Q_{\delta_k}$ . Since  $\delta_k \rightarrow \delta$ , we can use (38) to deduce

$$\begin{aligned} C(\delta_k - \delta) &\geq \inf(P_\delta) - \inf(P_{\delta_k}) \geq J(\bar{u}) - J(u_k) \\ &> q_k d_{Q_\delta}(G(y_k)) = q_k(\delta_k - \delta) \quad \forall k \geq k_\epsilon, \end{aligned}$$

which is not possible.  $\square$

Since  $J_q$  is not Gâteaux differentiable on  $Q_\delta$ , we are going to modify slightly this functional to attain the differentiability necessary for the proof.

**Proposition 3.** *Let us take  $q \geq q_0$  and for every  $\epsilon > 0$  let us consider the problem*

$$(P_{\delta,\epsilon}) \quad \inf_{u \in \mathcal{U}} J_{q,\epsilon}(u) = J(u) + q \{d_{Q_\delta}(G(y_u))^2 + \epsilon^2\}^{1/2}.$$

*Then  $\inf(P_{\delta,\epsilon}) \rightarrow \inf(P_\delta)$  when  $\epsilon \rightarrow 0$ .*

*Proof.* It is an immediate consequence of the inequality

$$J_q(u) \leq J_{q,\epsilon}(u) \leq J_q(u) + q\epsilon \quad \forall u \in \mathcal{U}. \quad \square$$

Finally we are ready to prove the strong Pontryagin's principle.

**Theorem 3.** *If  $(P_\delta)$  is strongly stable and  $\bar{u}$  is a solution of this problem, then Theorem 2 remains to be true with  $\bar{\alpha} = 1$ .*

*Proof.* Propositions 2 and 3 imply that  $u$  is a  $\sigma_\epsilon^2$  solution of  $(P_{\delta,\epsilon})$ , with  $\sigma_\epsilon \rightarrow 0$  when  $\epsilon \rightarrow 0$ , i.e.

$$J_{q,\epsilon}(u) \leq \inf(P_{\delta,\epsilon}) + \sigma_\epsilon^2.$$

Then we can apply again Ekeland's principle and deduce the existence of an element  $u' \in \mathcal{U}$  such that

$$d(u', \bar{u}) \leq \sigma_\epsilon, \quad J_{q,\epsilon}(u') \leq J_{q,\epsilon}(\bar{u}),$$

and

$$J_{q,\epsilon}(u') \leq J_{q,\epsilon}(u) + \sigma_\epsilon d(u', u) \quad \forall u \in \mathcal{U},$$

where once again  $d$  denotes the Ekeland's distance. Now we argue as in the proof of Theorem 2 and replace (31) by

$$-\sigma_\epsilon |\Omega| \leq \lim_{\rho \rightarrow 0} \frac{J_{q,\epsilon}(u_p^\epsilon) - J_{q,\epsilon}(u')}{\rho} = z^{0,\epsilon} + \langle \mu', DG(y')z' \rangle,$$

where  $\mu' \in Z'$  is given by

$$\mu' = \begin{cases} \frac{qd_{Q_\delta}(G(y'))}{\{d_{Q_\delta}(G(y'))^2 + \epsilon^2\}^{1/2}} \nabla d_{Q_\delta}(G(y')) & \text{if } G(y') \notin Q_\delta, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore we have  $\|\mu'\|_{Z'} \leq q$  for every  $\epsilon > 0$ . Now we can take a subsequence that converges weakly\* to an element  $\bar{\mu} \in Z'$ . The rest is as in the proof of Theorem 2, taking  $\alpha_\epsilon = 1$ .  $\square$

## References

1. R.A. Adams, *Sobolev spaces*, Academic Press, New York, 1975.
2. J.F. Bonnans, *Pontryagin's principle for the optimal control of semilinear elliptic systems with state constraints*, 30th IEEE Conference on Control and Decision (Brighton, England), 1991, pp. 1976-1979.
3. J.F. Bonnans and E. Casas, *Un principe de Pontryagin pour le contrôle des systèmes elliptiques*, J. Differential Equations **90** (1991), no. 2, 288-303.
4. ———, *A boundary Pontryagin's principle for the optimal control of state-constrained elliptic equations*, Internat. Ser. Numer. Math. **107** (1992), 241-249.
5. ———, *An extension of Pontryagin's principle for state-constrained optimal control of semilinear elliptic equations and variational inequalities*, SIAM J. Control Optim. (To appear).
6. J.F. Bonnans and D. Tiba, *Pontryagin's principle in the control of semilinear elliptic variational equations*, Appl. Math. Optim. **23** (1991), 299-312.
7. E. Casas, *Boundary control of semilinear elliptic equations with pointwise state constraints*, SIAM J. Control Optim. **31** (1993), no. 4, 993-1006.
8. ———, *Finite element approximations for some state-constrained optimal control problems*, Mathematics of the Analysis and Design of Process Control (P. Borne, S.G. Tzafestas and N.E. Radhy, eds.), North-Holland, Amsterdam, 1992.
9. E. Casas and J. Yong, *Maximum principle for state constrained optimal control problems governed by quasilinear elliptic equations*, Differential Integral Equations (To appear).
10. P.G. Ciarlet, *The finite element method for elliptic problems*, North-Holland, Amsterdam, 1978.
11. F.H. Clarke, *Optimization and nonsmooth analysis*, John Wiley & Sons, Toronto, 1983.
12. B.E.J. Dahlberg,  *$L^q$  estimates for Green potentials in Lipschitz domains*, Math. Scand. **44** (1979), no. 1, 149-170.
13. I. Ekeland, *Nonconvex minimization problems*, Bull. Amer. Math. Soc. **1** (1979), no. 3, 76-91.
14. H.O. Fattorini and H. Frankowska, *Necessary conditions for infinite dimensional control problems*, Math. Control Signals Systems **4** (1991), 41-67.
15. X. Li and Y. Yao, *Maximum principle of distributed parameter systems with time lags*, Distributed Parameter Systems (New York), Springer-Verlag, 1985, Lecture Notes in Control and Information Sciences 75, pp. 410-427.
16. X. Li and J. Yong, *Necessary conditions of optimal control for distributed parameter systems*, SIAM J. Control Optim. **29** (1991), 1203-1219.
17. J.B. Morrey, *Multiple integrals in the calculus of variations*, Springer-Verlag, New York, 1966.
18. J. Yong, *Pontryagin maximum principle for semilinear second order elliptic partial differential equations and variational inequalities with state constraints*, Differential Integral Equations (To appear).

Eduardo Casas  
Departamento de Matemática Aplicada y Ciencias de la Computación,  
E.T.S.I. Caminos, C. y P.,  
Universidad de Cantabria,  
Av. Los Castros s/n.,  
39071 Santander, Spain  
E-mail: [casas@masun1.unican.es](mailto:casas@masun1.unican.es)

# INVARIANCE OF THE HAMILTONIAN IN CONTROL PROBLEMS FOR SEMILINEAR PARABOLIC DISTRIBUTED PARAMETER SYSTEMS

H. O. FATTORINI

Department of Mathematics  
University of California

**ABSTRACT.** We consider a general optimal control problem for a distributed parameter described by a semilinear parabolic equation. We show that the Hamiltonian is invariant over optimally controlled trajectories, and that it vanishes when the terminal time is free. The method is that of approximation of the original control system by a sequence of "smoothed" control systems.

1991 *Mathematics Subject Classification.* 93E20, 93E25

*Key words and phrases.* Lagrange multiplier rule, Kuhn-Tucker conditions, maximum principle, optimal control.

## 1. Introduction

Let  $E$  be a Banach space,  $A$  the infinitesimal generator of a strongly continuous semigroup  $S(t)$  in  $E$ . Consider the optimal control problem of minimizing the functional

$$J_0(t) = \int_0^t f_0(y(\tau), u(\tau)) d\tau$$

among all trajectories of the autonomous semilinear differential system in  $E$

$$y'(t) = Ay(t) + f(y(t), u(t)), \quad y(0) = \zeta,$$

subject to a control constraint  $u(t) \in U(0 \leq t \leq \bar{t})$  and a target condition  $y(\bar{t}) \in Y$ . In situations where it can be proved, Pontryagin's maximum principle has two statements: one is that the Hamiltonian

$$H(\bar{y}(t), z_0, \bar{z}(t), v) = z_0 f_0(\bar{y}(t), v) + \langle \bar{z}(t), A\bar{y}(t) \rangle + \langle \bar{z}(t), f(\bar{y}(t), v) \rangle$$

---

This work was supported in part by the National Science Foundation under grant DMS-9221819.



is maximized over  $U$  by  $v = \bar{u}(t)$  along the trajectory, where  $\bar{u}(\cdot)$  is the optimal control,  $\bar{y}(t)$  the optimal trajectory and the costate  $z(t)$  solves the adjoint variational equation

$$\dot{\bar{z}}(t) = -\{A^* + \partial_y f(\bar{y}(t), \bar{u}(t))^*\} \bar{z}(t) - z_0 \partial_y f_0(\bar{y}(t), \bar{u}(t)), \quad \bar{z}(\bar{t}) = z$$

in  $0 \leq t \leq \bar{t}$ ;  $\partial_y$  indicates Fréchet derivative with respect to  $y$  and  $(z_0, z) \neq 0$ . The other statement is

$$H(\bar{y}(t), z_0, \bar{z}(t), \bar{u}(t)) = c \quad (0 \leq t \leq \bar{t})$$

with  $c = 0$  if the terminal time  $\bar{t}$  is free. In both statements we face the problem that the Hamiltonian (precisely, the term  $\langle \bar{z}(t), A\bar{y}(t) \rangle$ ) may not be defined: there are situations where  $\bar{y}(t)$  (resp.  $\bar{z}(t)$ ) does not belong to the domain of any fractional power of  $A$  (resp. of  $A^*$ ). This is not important for the first statement, since  $\langle \bar{z}(t), A\bar{y}(t) \rangle$  does not depend on  $v$  and can be dropped. However, this is not the case for the second statement, and the only results of this sort we know (coming from the maximum principle rather than from the dynamic programming approach) are those in [3]. They are very restrictive on nonlinearities; for instance the most general, if applied to a semilinear heat equation

$$(1.1) \quad y_t(t, x) = \Delta y(t, x) + f(y(t, x)) + u(t, x)$$

in  $L^2(\Omega)$  ( $\Omega$  a domain  $\subseteq \mathbb{R}^m$ ) would require  $L^2$ -continuity of  $f$ , thus preempting (for instance) any powers of  $y$ .

We present in this paper a new approach to invariance of the Hamiltonian based on another of the results in [3] and on an approximation process. It is not more or less general than the ones in [3], but applies for instance to (1.1) in the space  $C(\bar{\Omega})$  of continuous functions in  $\bar{\Omega}$ , and is very tolerant of nonlinearities; for instance, we may take as  $f$  in (1.1) a polynomial or arbitrarily high order or even a function of faster growth as long as certain dissipativity conditions that prevent finite time blowup of solutions are satisfied (see §2). See §6 for comparison with [3] and for possible generalizations.

## 2. The control system

Instead of the Laplacian, we work with

$$(2.1) \quad Ay = \sum_{j=1}^m \sum_{k=1}^m \partial^j (a_{jk}(x) \partial^k y) + \sum_{j=1}^m b_j(x) \partial^j y + c(x)y,$$

in a bounded domain  $\Omega \subseteq \mathbb{R}^m$  with boundary  $\Gamma$  of class  $C^{(2)}$ . The leading coefficients  $a_{jk}$  are continuously differentiable in  $\bar{\Omega}$ ,  $a_{jk} = a_{kj}$  and  $\sum \sum a_{jk}(x) \xi_j \xi_k \geq \kappa \|\xi\|^2$  ( $x \in \Omega$ ,  $\xi \in \mathbb{R}^m$ ); the  $b_j$  and  $c$  are continuous in  $\bar{\Omega}$ . We indicated by  $\partial$  either the Dirichlet boundary condition  $y = 0$  ( $x \in \Gamma$ ) or a variational boundary condition

$$\partial^\nu y(x) = \gamma(x)y(x) \quad (x \in \Gamma),$$

where  $\partial^\nu$  is the *conormal derivative*, i.e. the derivative in the direction of the conormal vector  $\{\nu_j(x)\} = \{\sum a_{jk}(x)\nu_k(x)\}$ .  $\{\nu_k(x)\}$  the outer normal vector on  $\Gamma$ . The coefficient  $\gamma(x)$  is assumed continuous in  $\Gamma$ . In general,  $A(\beta)$  is  $A$  with domain restricted by a boundary condition  $\beta$ .

The *adjoint*  $A'$  of (2.1) is

$$A'y = \sum_{j=1}^m \sum_{k=1}^m \partial^j (a_{jk}(x) \partial^k y) - \sum_{j=1}^m \partial_j (b_j(x) y) + c(x) y;$$

and the *adjoint boundary condition*  $\beta'$  is  $\beta' = \beta$  for the Dirichlet boundary condition; for a variational boundary condition  $\beta'$  is

$$\partial^\nu y(x) = (\gamma(x) + b(x))y(x)$$

with  $b(x) = \sum b_j(x)\nu_j(x)$ .

Denote by  $C(\bar{\Omega})$  the space of continuous functions  $y(\cdot)$  in  $\bar{\Omega}$  endowed with the supremum norm;  $C_0(\bar{\Omega})$  is the subspace defined by  $y = 0$  on  $\Gamma$ . The dual  $C(\bar{\Omega})^*$  of  $C(\bar{\Omega})$  can be identified with the space  $\Sigma(\bar{\Omega})$  of regular, bounded Borel measures  $\mu$  in  $\bar{\Omega}$  with the total variation norm; the duality pairing is  $\langle \mu, f \rangle = \int f(x)\mu(dx)$ . The dual  $C_0(\bar{\Omega})$  can be identified with the subspace of  $\Sigma(\bar{\Omega})$  whose elements vanish in  $\Gamma$ . We have the isometric imbeddings  $C(\bar{\Omega}) \subseteq L^\infty(\bar{\Omega})$ ,  $L^1(\bar{\Omega}) \subseteq \Sigma(\bar{\Omega})$ .

Let  $\beta$  be a variational boundary condition, and let  $A_c(\beta)$  be the operator in  $C(\bar{\Omega})$  defined as follows:

$$D(A_c(\beta)) = \left\{ y \in \bigcap_{p \geq 1} W^{2,p}(\Omega)_\beta; \quad Ay \in C(\bar{\Omega}) \right\}$$

and  $A_c(\beta)y = Ay$ , where  $W^{2,p}(\Omega)_\beta$  denotes the subspace of the Sobolev space  $W^{2,p}(\Omega)$  whose elements satisfy  $\beta$  on  $\Gamma$ . The operator  $A_c(\beta)$  generates a compact analytic semigroup  $S_c(t; A, \beta)$  in  $C(\bar{\Omega})$ . If  $\beta$  is the Dirichlet boundary condition the same is true in the space  $C_0(\bar{\Omega})$ ; in the definition of the domain, we require  $Ay \in C_0(\bar{\Omega})$ .

The dual semigroup  $S_c(t; A, \beta)^*$  in  $\Sigma(\bar{\Omega})(\Sigma_0(\bar{\Omega}))$  for the Dirichlet boundary condition) is compact and analytic in  $t > 0$  but not strongly continuous at  $t = 0$ ; however,  $S_c(t; A, \beta)^* \mu \rightarrow \mu$   $C(\bar{\Omega})$ -weakly in  $\Sigma(\bar{\Omega})$  ( $\Sigma_0(\bar{\Omega})$  for the Dirichlet boundary condition). The restriction of  $S_c(t; A, \beta)^*$  to  $L^1(\bar{\Omega})$  is a compact analytic semigroup  $S_1(t; A', \beta')$  ( $A'$  the adjoint of  $A$ ,  $\beta'$  the adjoint of  $\beta$ ) whose infinitesimal generator  $A'_1(\beta')$  is characterized as follows:  $D(A'_1(\beta'))$  consists of all elements  $y \in L^1(\bar{\Omega})$  such that there exists  $z (= A'_1(\beta')y)$  in  $L^1(\bar{\Omega})$  with

$$\int_{\Omega} y(x)(A(\beta)v)(x)dx = \int_{\Omega} z(x)v(x)dx$$

for every  $v$  in the subspace  $C^{(2)}(\bar{\Omega})_\beta$  of all  $y \in C^{(2)}(\bar{\Omega})$  that satisfy the boundary condition  $\beta$ . The dual semigroup  $S_1(t; A', \beta')^*$  is compact and analytic in  $t > 0$ , but not strongly continuous at  $t \rightarrow 0$ , although has the same weak continuity properties

as  $S_r(t; A, \beta)^*$ . Finally,  $S_r(t; A, \beta)$  is the restriction of  $S_1(t; A', \beta')^*$  to  $C(\bar{\Omega})$  ( $C_0(\bar{\Omega})$  for the Dirichlet boundary condition). Restriction relations also hold for the generators:  $A'_1(\beta')$  is the restriction of  $A_r(\beta)^*$  to the domain  $D(A'_1(\beta')) = \{y \in L^1(\Omega); A_r(\beta)^* y \in L^1(\Omega)\}$  and  $A_r(\beta)$  is the restriction of  $A'_1(\beta')^*$  to  $D(A_r(\beta)) = \{y \in C(\bar{\Omega}); A'_1(\beta')^* y \in C(\bar{\Omega})\}$  ( $C_0(\bar{\Omega})$  for the Dirichlet boundary condition).

We fix below the operator  $A$  and the boundary condition  $\beta$  and use the shorthand  $S_r(t) = S_r(t; A, \beta)$ ,  $S_\infty(t) = S_1(t; A', \beta')^*$ ,  $S'_1(t) = S_1(t; A', \beta')$ ,  $S'_\Sigma(t) = S_r(t; A, \beta)^*$ , so that  $S_r(t)$  (resp.  $S'_1(t)$ ) is the restriction of  $S_\infty(t)$  to  $C(\bar{\Omega})$  (resp. of  $S'_\Sigma(t)$  to  $L^1(\Omega)$ ). Likewise, we write  $A_r = A_r(\beta)$ ,  $A'_1 = A'_1(\beta')$ . Also, we leave to the reader to add " $C_0(\bar{\Omega})$  for the Dirichlet boundary condition" or " $\Sigma_0(\bar{\Omega})$  for the Dirichlet boundary condition" whenever necessary.

The control system is

$$(2.2) \quad \mathbf{y}'(t) = A_r \mathbf{y}(t) + \mathbf{f}(\mathbf{y}(t)) + \mathbf{u}(t), \quad \mathbf{y}(0) = \zeta$$

and is modelled on (1.1), so that the boldface letters mean:  $\mathbf{y}(t)(x) = y(t, x)$ , and  $\mathbf{u}(t)(x) = u(t, x)$ . Instantaneous values of  $\mathbf{y}$  are in  $C(\bar{\Omega})$ . We want to admit  $L^\infty((0, T) \times \Omega)$  as control space, thus we cannot take  $\mathbf{u}(t) \in C(\bar{\Omega})$ . To find the "right" control space we take an arbitrary separable Banach space  $X$  and call  $L_w^\infty(0, T; X^*)$  the space of all  $X$ -weakly measurable  $X^*$ -valued functions  $g(\cdot)$  endowed with the essential supremum norm (separability of  $X$  implies measurability of the norm). In this notation,

$$(2.3) \quad L^\infty((0, T) \times \Omega) = L_w^\infty(0, T; L^\infty(\Omega))$$

thus we assume  $\mathbf{u}(\cdot) \in L_w^\infty(0, T; L^\infty(\Omega))$  in (2.2).

By definition, solutions of (2.2) are  $C(\bar{\Omega})$ -valued continuous solutions of the associated integral equation

$$(2.4) \quad y(t) = S_r(t)\zeta + \int_0^t S_r(t-\tau)\mathbf{f}(\mathbf{y}(\tau))d\tau + \int_0^t S_\infty(t-\tau)\mathbf{u}(\tau)d\tau$$

where we use  $S_\infty(t) \supseteq S_r(t)$  since  $\mathbf{u}(\cdot)$  takes values in  $L^\infty(\Omega)$ . Although  $\mathbf{u}(\cdot)$  is merely weakly measurable,  $\tau \rightarrow S_\infty(t-\tau)\mathbf{u}(\tau)$  is strongly measurable. ([6, Lemma 4.1]) thus the last integral is a Bochner integral and produces a continuous  $C(\bar{\Omega})$ -valued function. The operator  $\mathbf{f} : C(\bar{\Omega}) \rightarrow C(\bar{\Omega})$  in (2.2) and (2.4) is  $\mathbf{f}(y)(x) = f(y(x))$ . If  $f(y)$  is locally Lipschitz continuous with respect to  $y$  then  $\mathbf{f}$  is locally bounded and locally Lipschitz continuous and this is enough to insure local existence and uniqueness of solutions of (2.4); of course, we require  $f(0) = 0$  for the Dirichlet boundary condition. We shall call  $\mathbf{y}(t, \mathbf{u})$  (or  $\mathbf{y}(t, \zeta, \mathbf{u})$  if  $\zeta$  is not fixed) the solution of (2.4) for  $\mathbf{u}(\cdot) \in L_w^\infty(0, T; L^\infty(\Omega))$ .

We say that (2.2) (or any control system) has the *global existence property* (GEP) in an interval  $0 \leq t \leq \bar{t}$  if for every  $C, K > 0$  there exists  $L = L(C, K) > 0$  such that if  $\|\zeta\| \leq C$  and  $\|\mathbf{u}(\cdot)\| \leq K$  then  $\mathbf{y}(t, \zeta, \mathbf{u})$  exists in  $0 \leq t \leq \bar{t}$  and  $\|\mathbf{y}(t, \zeta, \mathbf{u})\| \leq L$ . One of the many conditions that produce the GEP is

**Lemma 2.1.** *Assume that*

$$(2.5) \quad \langle \mu, f(y) \rangle \leq c(1 + \|y\|^2) \quad (y \in C(\bar{\Omega}), \mu \in \Theta(y)),$$

where the duality set  $\Theta(y) \subseteq \Sigma(\bar{\Omega})$  of an element  $y \in C(\bar{\Omega})$  is the set of all  $\mu \in \Sigma(\bar{\Omega})$  such that  $\langle \mu, y \rangle = \|\mu\|^2 = \|y\|^2$ . Then (2.2) has the GEP.

To see what (2.5) means for, say, the Dirichlet boundary condition, note that if  $y(\cdot) \in C_0(\bar{\Omega})$  then  $\Theta(y)$  consists of all  $\mu \in \Sigma(\bar{\Omega})$  supported by  $e = \{x \in \bar{\Omega}; |y(x)| = \|y\|\}$  and such that  $\|\mu\| = \|y\|$ ,  $\mu \geq 0$  in  $e_+ = \{x \in \bar{\Omega}; y(x) = \|y(x)\|\}$ ,  $\mu \leq 0$  in  $e_- = \{x \in \bar{\Omega}; y(x) = -\|y(x)\|\}$ . Accordingly, if

$$(2.6) \quad (\operatorname{sgn} y)f(y) \leq C(1 + |y|) \quad (-\infty < y < \infty)$$

then

$$\begin{aligned} \langle \mu, f(y(\cdot)) \rangle &= \int_e f(y(x))\mu(dx) = \int_e \operatorname{sgn} y(x)f(y(x))|\mu|(dx) \\ &= c \int_e (1 + |y(x)|)|\mu|(dx) = c(\|y\| + \|y\|^2) \leq 2c(1 + \|y\|^2), \end{aligned}$$

so that (2.5) holds. Condition (2.6) is easily seen to be satisfied by any polynomial containing only odd powers with negative coefficients, for instance  $f(y) = -y^{2k+1}$ .

### 3. Approximation of the state

The scheme uses

$$(3.1.n) \quad \mathbf{y}'_n(t) = A_c \mathbf{y}_n(t) + S_c(\varepsilon_n) \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}_n(t)) + S_\infty(\varepsilon_n) \mathbf{u}(t), \quad \mathbf{y}_n(0) = \zeta$$

with  $\varepsilon_n > 0$ ,  $\varepsilon_n \rightarrow 0$ . As for (2.2),  $\mathbf{y}_n(t, \zeta, \mathbf{u})$  indicates dependence on  $\zeta, \mathbf{u}$ . We assume that  $\|S_c(t)\| \leq Ce^{-\omega t}$  with  $\omega > 0$  (which can always be achieved by a translation) so that fractional powers  $(-A_c)^\alpha$  can be defined; since  $S_c(t)$  is analytic, if  $\alpha \geq 0$ ,

$$(3.2) \quad \|(-A_c)^\alpha S_c(t)\| \leq C_\alpha t^{-\alpha} \quad (t \geq 0).$$

**Lemma 3.1.** *Let  $0 \leq \alpha < 1$ . Then the operator*

$$(3.3) \quad \Lambda_\alpha g(\cdot) = \int_0^t (-A_c)^\alpha S_\infty(t - \sigma) \mathbf{u}(\sigma) d\sigma,$$

is compact from  $L^\infty_w(0, T; L^\infty(\Omega))$  into  $C(0, T; C(\bar{\Omega}))$ .

The proof is basically the same as that of its abstract counterpart, Theorem 3.1 (or, rather, Theorem 9.1) in [7].

We use below the fact that  $L^\infty_w(0, T; L^\infty(\Omega)) = L^1(0, T; L^1(\Omega))^*$  (which is obvious by (2.3)). We recall that  $L^1(0, T; L^1(\Omega))$ -weak convergence of a sequence implies boundedness; also, a  $L^1(0, T; L^1(\Omega))$ -weakly compact set has to be bounded.

**Corollary 3.2.** Let  $\mathbf{u}(\cdot) \in L_w^\infty(0, \bar{t}; L^\infty(\Omega))$  be such that  $\mathbf{y}(t, \mathbf{u})$  exists in  $0 \leq t \leq \bar{t}$ ,  $\{\zeta_n\} \subset C(\bar{\Omega})$  with  $\zeta_n \rightarrow \zeta$ ,  $\{\mathbf{u}_n(\cdot)\} \subseteq L_w^\infty(0, \bar{t}; L^\infty(\Omega))$  with  $\mathbf{u}_n(\cdot) \rightarrow \mathbf{u}(\cdot)$   $L^1(0, \bar{t}; L^1(\Omega))$ -weakly. Then, if  $\varepsilon > 0$  there exists  $n = n_\varepsilon$  such that  $\mathbf{y}_n(t, \zeta_n, \mathbf{u}_n)$  exists in  $0 \leq t \leq \bar{t}$  for  $n \geq n_\varepsilon$  and

$$(3.4) \quad \|\mathbf{y}_n(t, \zeta_n, \mathbf{u}_n) - \mathbf{y}(t, \zeta, \mathbf{u})\| \leq \varepsilon$$

$$(3.5) \quad \|(-A_c)^\alpha \mathbf{y}_n(t, \zeta_n, \mathbf{u}_n)\|, \|(-A_c)^\alpha \mathbf{y}(t, \zeta, \mathbf{u})\| \leq C_\alpha t^{-\alpha}$$

$$(3.6) \quad \|(-A_c)^\alpha \mathbf{y}_n(t, \zeta_n, \mathbf{u}_n) - (-A_c)^\alpha \mathbf{y}(t, \zeta, \mathbf{u})\| \rightarrow 0$$

uniformly in  $\delta \leq t \leq \bar{t}$  for any  $\delta > 0$  (in  $0 \leq t \leq \bar{t}$  if  $\alpha = 0$ ).

*Proof:* Denote by  $[0, t_n]$  the maximum subinterval of  $[0, \bar{t}]$  where  $\mathbf{y}_n(t, \zeta_n, \mathbf{u}_n)$  exists and satisfies (3.4). In this interval, we have

$$(3.7) \quad \begin{aligned} & (-A_c)^\alpha (\mathbf{y}_n(t, \zeta_n, \mathbf{u}_n) - \mathbf{y}(t, \zeta, \mathbf{u})) = (-A_c)^\alpha S_c(t)(\zeta_n - \zeta) \\ & + \int_0^t (-A_c)^\alpha S_c(t - \tau + \varepsilon_n) \{ \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}_n(\tau, \zeta_n, \mathbf{u}_n)) \\ & \quad - \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}(\tau, \zeta, \mathbf{u})) \} d\tau \\ & + \int_0^t (-A_c)^\alpha S_c(t - \tau + \varepsilon_n) \{ \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}(\tau, \zeta, \mathbf{u})) \\ & \quad - \mathbf{f}(\mathbf{y}(\tau, \zeta, \mathbf{u})) \} d\tau \\ & + \int_0^t (-A_c)^\alpha (S_c(t - \tau + \varepsilon_n) - S_c(t - \tau)) \mathbf{f}(\mathbf{y}(\tau, \zeta, \mathbf{u})) d\tau \\ & + \int_0^t (-A_c)^\alpha (S_\infty(t - \tau + \varepsilon_n) - S_\infty(t - \tau)) \mathbf{u}_n(\tau) d\tau \\ & + \int_0^t (-A_c)^\alpha S_\infty(t - \tau) (\mathbf{u}_n(\tau) - \mathbf{u}(\tau)) d\tau \\ & = \int_0^t (-A_c)^\alpha S_c(t - \tau + \varepsilon_n) \{ \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}_n(\tau, \zeta_n, \mathbf{u}_n)) \\ & \quad - \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}(\tau, \zeta, \mathbf{u})) \} d\tau + \kappa_n(t) \end{aligned}$$

which we first estimate for  $\alpha = 0$ . In the first three integrals making up  $\kappa_n(t)$  we break up the interval of integration into  $(0, \rho t)$  and  $(\rho t, t)$  with  $0 < \rho < 1$ . In the first interval, we use uniform continuity of  $S_c(t)$  and  $S_\infty(t)$  in  $t > 0$ ; in the second the bound for  $\mathbf{u}_n(t)$ . The three integrals are shown in this way to tend to zero uniformly in  $0 \leq t \leq \bar{t}$ . The same is true of the fourth using Lemma 3.1. We estimate by Gronwall's inequality and, taking  $n \geq a$  certain  $n_\varepsilon$  we deduce that  $\|\mathbf{y}_n(t, \zeta_n, \mathbf{u}_n) - \mathbf{y}(t, \zeta, \mathbf{u})\| \leq \varepsilon/2$  and contradict the maximality of  $[0, t_n]$  unless  $t_n = \bar{t}$ .

The estimations for  $\alpha > 0$  use the *generalized Gronwall inequality* [11, p. 188 Lemma 7.1.1]; if  $b \leq 0$ ,  $\gamma > -1$  and  $a(t)$ ,  $u(t)$  are nonnegative and locally integrable in  $0 \leq t \leq T$  with

$$u(t) \leq a(t) + b \int_0^t (t-\sigma)^\gamma u(\sigma) d\sigma \quad (0 \leq t \leq T),$$

then there exists  $C$  depending only on  $\gamma, T$  such that

$$u(t) \leq a(t) + Cb \int_0^t (t-\sigma)^\gamma a(\sigma) d\sigma \quad (0 \leq t \leq T).$$

The treatment of all integrals is the same and we omit the details. Note that the singular behavior at  $t = 0$  comes only from the first term  $(-A_c)^\alpha (\zeta_n - \zeta)$  of  $\kappa_n(t)$ , not from the integral terms.

**Corollary 3.3.** *Assume (2.2) has the GEP in  $0 \leq t \leq \bar{t}$ . Then (3.1.n) has the GEP in  $0 \leq t \leq \bar{t}$  for  $n$  large enough, with constants  $C, K, L(C, K) + \varepsilon$ ,  $\varepsilon > 0$ .*

*Proof:* Let  $\|\zeta\| \leq C$ ,  $\|\mathbf{u}(\cdot)\| \leq K$ . Use Lemma 3.2 ( $\alpha = 0$ ) with  $\zeta_n = \zeta$ ,  $\mathbf{u}_n(\cdot) = \mathbf{u}(\cdot)$ . Notice that the estimation of the first three integrals making up  $\kappa_n(t)$  is independent of  $\mathbf{u}$  and that the fourth integral is absent. We then obtain that  $\mathbf{y}_n(t, \zeta, \mathbf{u})$  exists in  $0 \leq t \leq \bar{t}$  and satisfies  $\|\mathbf{y}_n(t, \zeta, \mathbf{u})\| \leq L + \varepsilon$ . Compactness of  $S_c(t)$  (or Lemma 3.1) is not used here.

We assume from now on that (2.2) has the GEP. The proof of the following result is very similar to that of Corollary 3.3 and is thus omitted.

**Corollary 3.4.** *Let  $0 \leq \alpha < 1$ .  $\{\mathbf{u}_k(\cdot)\}$  a  $L^1(0, t; L^1(\Omega))$ -weakly convergent sequence in  $L_w^\infty(0, t; L^\infty(\Omega))$ . Then*

$$(3.8) \quad \|(-A_c)^\alpha \mathbf{y}(t, \zeta, \mathbf{u}_k) - (-A_c)^\alpha \mathbf{y}(t, \zeta, \mathbf{u})\| \rightarrow 0 \quad \text{as } k \rightarrow \infty$$

$$(3.9) \quad \|(-A_c)^\alpha \mathbf{y}_n(t, \zeta, \mathbf{u}_k) - (-A_c)^\alpha \mathbf{y}_n(t, \zeta, \mathbf{u})\| \rightarrow 0 \quad \text{as } k \rightarrow \infty$$

*uniformly in  $0 \leq t \leq \bar{t}$ .*

A *cost functional* for (2.2) is a real valued function  $y_0(t, \mathbf{y}, \mathbf{u})$  defined in  $t > 0$  and for each  $t$ , for  $\mathbf{y} = \mathbf{y}(\cdot) \in C(0, t; C(\bar{\Omega}))$  and  $\mathbf{u} = \mathbf{u}(\cdot) \in L_w^\infty(0, t; L^\infty(\Omega))$ . It is *weakly lower semicontinuous* (WLS) if

$$(3.10) \quad y_0(t, \bar{\mathbf{y}}, \bar{\mathbf{u}}) \leq \limsup_{k \rightarrow \infty} y_0(t_k, \mathbf{y}_k, \mathbf{u}_k)$$

for every sequence  $\{t_k\} \subseteq \mathbb{R}$  with  $t_k \rightarrow \bar{t}$ , every sequence  $\{\mathbf{y}_k(\cdot)\} \subseteq C(0, t_k; C(\bar{\Omega}))$  with  $\mathbf{y}(\cdot) \rightarrow \bar{\mathbf{y}}(\cdot)$  in  $C(0, T; C(\bar{\Omega}))$  and every sequence  $\{\mathbf{u}_k(\cdot)\} \subseteq L_w^\infty(0, T; L^\infty(\Omega))$  with  $\mathbf{u}_k(\cdot) \rightarrow \bar{\mathbf{u}}(\cdot)$   $L^1(0, T; L^1(\Omega))$ -weakly. (Here  $T > \bar{t}$  and the passage from spaces in  $[0, t_n]$  to spaces in  $[0, T]$  involves obvious extensions.)

We equip the control system (2.2) with a WLS cost functional  $y_0(t, \mathbf{y}(\zeta, \mathbf{u}), \mathbf{u})$  and a control set  $U \subseteq L^\infty(\Omega)$ ; the admissible control space  $C_{ad}(0, T; U)$  consists of all  $\mathbf{u}(\cdot) \in L_w^\infty(0, t; L^\infty(\Omega))$  such that  $\mathbf{u}(t) \in U$  a.e. The optimal control problem includes a target condition  $\mathbf{y}(\bar{t}, \zeta, \mathbf{u}) = Y = \text{target set} \subseteq C(\bar{\Omega})$ , and we assume  $Y$  closed. Call  $m$  the minimum of  $y_0(t, \mathbf{y}(\zeta, \mathbf{u}), \mathbf{u})$  in  $C_{ad}(0, \bar{t}; U)$  subject to the target condition. Assuming that  $-\infty < m < \infty$ , a sequence  $\{\mathbf{u}^k(\cdot)\}$ ,  $\mathbf{u}^k(\cdot) \in C_{ad}(0, t_k; U)$  is a minimizing sequence if we have

$$(3.11) \quad \limsup_{k \rightarrow \infty} y_0(t_k, \mathbf{y}(\zeta, \mathbf{u}^k), \mathbf{u}^k) \leq m, \quad \lim_{k \rightarrow \infty} \text{dist}(\mathbf{y}(t_k, \zeta, \mathbf{u}^k), Y) \rightarrow 0.$$

The definitions are the same for (3.1.n).

**Theorem 3.5.** Let  $C_{ad}(0, T; U)$  be  $L^1(0, T; L^1(\Omega))$ -weakly compact in  $L_w^\infty(0, T; L^\infty(\Omega))$ , and assume  $y_0(t, \mathbf{y}, \mathbf{u})$  is weakly lower semicontinuous. Let  $\{\mathbf{u}^k(\cdot)\}$ ,  $\mathbf{u}^k(\cdot) \in C_{ad}(0, t_k; U)$  be a minimizing sequence with  $t_k \rightarrow \bar{t} < \infty$ . Then (a generalized sequence of)  $\{\mathbf{u}^k(\cdot)\}$  converges  $L^1(0, T; L^1(\Omega))$ -weakly to an optimal control  $\bar{\mathbf{u}}(\cdot)$  for (2.2). The same result holds for (3.1.n).

*Proof:* By the GEP we know that  $\{\mathbf{y}(t, \mathbf{u}^k)\}$  is bounded, so that using Alaoglu's theorem in  $L_w^\infty(0, T; L^\infty(\Omega))$  we may select a subsequence (denoted equally  $\{\mathbf{u}^k(\cdot)\}$ ) such that both  $\{\mathbf{u}^k(\cdot)\}$  and  $\{\mathbf{f}(\mathbf{y}(\cdot, \mathbf{u}^k))\}$  are  $L^1(0, T; L^1(\Omega))$ -weakly convergent. Then, using Corollary 3.4 we deduce that  $\{\mathbf{y}(t, \mathbf{u}^k)\}$  is convergent in  $C(0, T; C(\bar{\Omega}))$ . Some acrobatics are necessary to take care of the moving  $t_n$ . For a very similar argument see for instance [9, Theorem 4.4]. The reasoning for (3.1.n) is of course the same.

We note that if  $\mathbf{u}(\cdot) \in C_{ad}(0, \bar{t}; U)$  is such that  $\mathbf{y}(\bar{t}, \zeta, \mathbf{u}) \in Y$  then a minimizing sequence for (2.2) automatically exists, although not necessarily with  $\{t_k\}$  bounded. The same result holds for (3.1.n).

A popular cost functional for (1.2) is

$$(3.12) \quad y_0(t, \mathbf{y}, \mathbf{u}) = \int_0^t \int_\Omega f_0(y(\tau, x, \zeta, u), u(\tau, x)) dx d\tau$$

where  $u(t, x) = \mathbf{u}(t)(x)$ ,  $y(t, x, \zeta, u) = \mathbf{y}(t, \zeta, \mathbf{u})(x)$ . The proof of the following result is standard and we omit it.

**Lemma 3.6.** Let  $f_0(y, u)$  be continuous in  $\bar{\Omega} \times \mathbb{R}$  and convex in  $u$  for  $y$  fixed. Then  $y_0(t, \mathbf{u})$  is weakly lower semicontinuous.

#### 4. Approximation of the costate

We assume that  $f$  is continuously differentiable, so that  $\mathbf{f}$  has a Fréchet derivative  $\partial \mathbf{f}$  with respect to  $y$  in  $C(\bar{\Omega})$  given by  $(\partial \mathbf{f}(y)z)(x) = f'(y(x))z(x)$ . The derivative is continuous with respect to  $y$  in the space  $(C(\bar{\Omega}), C(\bar{\Omega}))$  of bounded

operators from  $C(\bar{\Omega})$  into itself. The *adjoint variational equation* corresponding to (2.2) is

$$(4.1) \quad \mathbf{z}'(s) = -\{A'_1 + \partial \mathbf{f}(\mathbf{y}(s))^*\} \mathbf{z}(s) - \mathbf{g}(s), \quad \mathbf{z}(\bar{t}) = z$$

in  $L^1(\Omega)$ , to be solved backwards in  $0 \leq t \leq \bar{t}$ ; here  $z \in \Sigma(\bar{\Omega})$ ,  $\mathbf{g}(\cdot) \in L_w^\infty(0, T; \Sigma(\bar{\Omega}))$  and  $\mathbf{y}(\cdot) \in C(0, \bar{t}; C(\bar{\Omega}))$ . The operator  $\partial \mathbf{f}(\mathbf{y}(t))^*$  is multiplication by  $f'(y(x))$  and is thus continuous in  $(\Sigma(\bar{\Omega}), \Sigma(\bar{\Omega}))$  and in  $(L^1(\Omega), L^1(\Omega))$ . The treatment of (4.1) parallels that of (2.2) with  $\Sigma(\bar{\Omega})$  taking the role of  $L^\infty(\Omega)$  and  $L^1(\Omega)$  taking the role of  $C(\bar{\Omega})$ . A solution of (4.1) is a continuous solution of the integral equation

$$(4.2) \quad \mathbf{z}(s) = S'_\Sigma(\bar{t} - s)z + \int_s^{\bar{t}} S'_1(\sigma - s) \partial \mathbf{f}(\mathbf{y}(\sigma))^* \mathbf{z}(s) + \int_s^{\bar{t}} S'_\Sigma(\sigma - s) \mathbf{g}(\sigma) d\sigma,$$

where the first integral is a Bochner integral. So is the second, on account of strong measurability of  $\sigma \rightarrow S'_\Sigma(\sigma - s) \mathbf{g}(\sigma)$  [6, Lemma 4.1]. Note that we use  $S'_1$  in the first integral since solutions belong to  $L^1(\Omega)$ . Due to linearity, (4.2) enjoys global existence.

Since  $\{S_c(\varepsilon_n) \partial \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}(s))^*\}^* = S_c(\varepsilon_n)^* \partial \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}(s))^* S_c(\varepsilon_n)^*$  and  $S'_1(t)$  is the restriction to  $L^1(\Omega)$  of  $S_c(t)^*$ , the adjoint variational equation corresponding to (3.1.n) is

$$(4.3) \quad \begin{aligned} \mathbf{z}'_n(s) &= -\{A'_1 + S'_1(\varepsilon_n) \partial \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}_n(s))^* S'_1(\varepsilon_n)\} \mathbf{z}_n(s) - \mathbf{g}_n(s), \\ \mathbf{z}_n(\bar{t}) &= z_n \end{aligned}$$

also to be solved backwards in  $0 \leq t \leq \bar{t}$  with  $z_n \in \Sigma(\bar{\Omega})$  and  $\mathbf{g}_n(\cdot) \in L_w^\infty(0, T; \Sigma(\bar{\Omega}))$ ; as in (4.1),  $\mathbf{y}_n(\cdot) \in C(0, \bar{t}; C(\bar{\Omega}))$ . The associated integral equation is the corresponding replica of (4.2).

**Lemma 4.1.** *Let  $0 \leq \alpha < 1$ . Then the operator*

$$(4.4) \quad \Lambda_\alpha g(\cdot) = \int_s^{\bar{t}} (-A'_1)^\alpha S'_\Sigma(\sigma - s) \mathbf{g}(\sigma) d\sigma,$$

*is compact from  $L_w^\infty(0, T; \Sigma(\bar{\Omega}))$  into  $C(0, T; L^1(\Omega))$ .*

The proof follows from the same abstract scheme as does that of Lemma 3.1 (Theorems 3.1 and 9.1 in [7]).

Below, we use the equality  $L^1(0, T; C(\bar{\Omega}))^* = L_w^\infty(0, T; \Sigma(\bar{\Omega}))$ . This is a particular case of the formula  $L^1(0, T; X)^* = L_w^\infty(0, T; X^*)$  valid for any Banach space  $X$ ; see [7] for references.

**Corollary 4.2.** *Let  $\{z_n\} \subseteq \Sigma(\bar{\Omega})$  be  $C(\bar{\Omega})$ -weakly convergent to  $z \in \Sigma(\bar{\Omega})$  and let  $\{\mathbf{g}_n(\cdot)\} \subseteq L_w^\infty(0, \bar{t}; \Sigma(\bar{\Omega}))$  be  $L^1(0, \bar{t}; C(\bar{\Omega}))$ -weakly convergent to  $\mathbf{g}(\cdot) \in L_w^\infty(0, \bar{t}; \Sigma(\bar{\Omega}))$ . Finally, let  $\mathbf{y}_n(\cdot) \rightarrow \mathbf{y}(\cdot)$  in  $C(0, \bar{t}; C(\bar{\Omega}))$ , and  $0 \leq \alpha < 1$ . Then*

$$(4.5) \quad \|(-A'_1)^\alpha \mathbf{z}_n(\cdot)\|, \quad \|(-A'_1)^\alpha \mathbf{z}(t)\| \leq C_\alpha t^{-\alpha}$$

$$(4.6) \quad \|(-A'_1)^\alpha \mathbf{z}_n(\cdot) - (-A'_1)^\alpha \mathbf{z}(t)\| \rightarrow 0$$



uniformly in  $0 \leq t \leq \bar{t} - \delta$  for every  $\delta > 0$ .

*Proof:* The proof is just the same as that of Corollary 3.2 with one minor difference; the nonintegral term on the right side of (3.5) is  $(-A_c)^n S_c(t)(\zeta_n - \zeta)$  with  $\{\zeta_n\} \subseteq C(\bar{\Omega})$ ,  $\zeta_n \rightarrow \zeta$  strongly, while here the corresponding term is  $(-A'_1)^n S'_\Sigma(\bar{t} - s)(z_n - z)$  with  $\{z_n\} \subseteq \Sigma(\bar{\Omega})$ ,  $z_n \rightarrow z$   $C(\bar{\Omega})$ -weakly; however compactness of  $S'_\Sigma(\bar{t} - s)$  makes this a moot point.

## 5. The maximum principle and invariance of the Hamiltonian

We consider a cost functional

$$(5.1) \quad y_0(t, \mathbf{y}(\zeta, \mathbf{u}), \mathbf{u}) = \int_0^t \mathbf{f}_0(\mathbf{y}(\tau, \zeta, \mathbf{u}), \mathbf{u}(\tau)) d\tau$$

where  $f_0 : \bar{\Omega} \times L^\infty(\Omega) \rightarrow \mathbb{R}$  has a Fréchet differential  $\partial_y \mathbf{f}_0(y, u)$  with respect to  $y$ . We assume that  $\mathbf{f}_0(y, u)$  is  $C'(\bar{\Omega})$ -continuous in  $y$  for  $u$  fixed and that  $\partial_y \mathbf{f}_0(y, u)$  is  $C(\bar{\Omega})$ -continuous in  $y$  as a  $\Sigma(\bar{\Omega})$ -valued function for  $u$  fixed. Moreover, we require  $t \rightarrow f_0(y, \mathbf{u}(t))$  measurable and bounded and  $t \rightarrow \partial_y \mathbf{f}_0(y, \mathbf{u}(t))$   $C'(\bar{\Omega})$ -weakly measurable and bounded for every  $\mathbf{u}(\cdot) \in L^\infty_w(0, T; L^\infty(\Omega))$ . Finally, we assume that  $y_0(t, \mathbf{y}, \mathbf{u})$  is WLS and that

$$(5.2) \quad \mathbf{f}_0(y, u) \geq \delta > 0 \quad (y \in C'(\bar{\Omega}), u \in U).$$

For a functional (3.12), all of the above will hold if  $\mathbf{f}_0(y, u)$  is continuous in  $\bar{\Omega} \times \mathbb{R}$ , continuously differentiable with respect to  $y$  in  $\bar{\Omega} \times \mathbb{R}$  and positive. The system (2.2) is assumed to have the GEP, so that all the approximating systems (3.1.n) will also have the GEP for  $n$  large enough. We assume that the control set  $C_{ad}(0, T; U)$  is  $L^1(0, T; L^1(\Omega))$ -weakly closed in  $L^\infty_w(0, T; L^\infty(\Omega))$  and that the target set  $Y$  is closed.

Assume that a control  $\mathbf{u}(\cdot) \in C_{ad}(0, T; U)$  exists with  $\mathbf{y}(t, \zeta, \mathbf{u}) \in Y$  for some  $t$ . Then (see the remarks at the end of §3) a minimizing sequence for (2.2) exists. Condition (5.2) implies

$$(5.3) \quad y_0(t, \mathbf{y}(\zeta, \mathbf{u}), \mathbf{u}) \geq \delta t$$

thus for every minimizing sequence,  $\{t_n\}$  is automatically bounded. Hence, by Theorem 3.5, an optimal control  $\bar{\mathbf{v}}(\cdot)$  exists in an interval  $0 \leq t \leq \bar{t}$ .

We consider the approximating problems (3.1.n) with the same admissible control space  $C_{ad}(0, T; U)$  but a different initial condition  $\zeta_n$ . The cost functional is

$$(5.4) \quad y_{0n}(t, \mathbf{y}_n(\zeta, \mathbf{u}), \mathbf{u}) = \int_0^t \mathbf{f}_{0n}(\mathbf{y}_n(\tau, \zeta_n, \mathbf{u}), \mathbf{u}(\tau)) d\tau$$

where  $\mathbf{f}_{0n}(\mathbf{y}, \mathbf{u}) = \mathbf{f}_0(S_c(\varepsilon_n)y, u)$ . Obviously,  $\mathbf{f}_{0n}(y, u)$  satisfies (5.2) and has a Fréchet differential  $\partial_y \mathbf{f}_{0n}(y, u) = S_c(\varepsilon_n)^* \partial_y \mathbf{f}_0(S_c(\varepsilon_n)y, u)$  sharing all the relevant

properties of  $\partial_y \mathbf{f}_0(y, u)$ . We check easily that  $y_{on}$  is as well WLS. The target set  $Y_n$  is constructed as follows. As a particular case of Lemma 3.1,

$$\delta_n = \sup_{u \in C_{ad}(0, \bar{t}, U)} \|\mathbf{y}_n(\bar{t}, \zeta, u) - \mathbf{y}(\bar{t}, \zeta, u)\| \rightarrow 0.$$

We set  $Y_n = \{y; \text{dist}(y, Y) \leq 2\delta_n\}$  and choose  $\zeta_n \in D(A_c)$  tending to  $\zeta$  (in the norm of  $C(\bar{\Omega})$ ) in such a way that  $\mathbf{y}_n(\bar{t}, \zeta_n, \bar{\mathbf{v}}) \in Y_n$ . Accordingly, a minimizing sequence for (3.1.n) exists, thus an optimal control  $\bar{\mathbf{u}}^n(\cdot)$  exists in some interval  $[0, t_n]$  by Theorem 3.5. We have

$$\begin{aligned} & y_0(t_n, \mathbf{y}(\zeta, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n) \\ &= y_{on}(t_n, \mathbf{y}_n(\zeta_n, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n) + \{y_0(t_n, \mathbf{y}(\zeta, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n) - y_{on}(t_n, \mathbf{y}_n(\zeta_n, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n)\} \\ &\leq y_{on}(\bar{t}, \mathbf{y}_n(\zeta_n, \bar{\mathbf{v}}), \bar{\mathbf{v}}) + \{y_0(t_n, \mathbf{y}(\zeta, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n) - y_{on}(t_n, \mathbf{y}_n(\zeta_n, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n)\} \\ &\leq y_0(\bar{t}, \mathbf{y}(\zeta, \bar{\mathbf{v}}), \bar{\mathbf{v}}) + \{y_{on}(\bar{t}, \mathbf{y}_n(\zeta_n, \bar{\mathbf{v}}), \bar{\mathbf{v}}) - y_0(\bar{t}, \mathbf{y}(\zeta, \bar{\mathbf{v}}), \bar{\mathbf{v}})\} \\ &\quad + \{y_0(t_n, \mathbf{y}(\zeta, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n) - y_{on}(t_n, \mathbf{y}_n(\zeta_n, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n)\} \end{aligned}$$

with the two curly brackets tend to zero in view of Corollary 3.2. It then results that  $\{\bar{\mathbf{u}}^n\}$  is a minimizing sequence for the original problem (2.2). Applying once again Theorem 3.5, we deduce that a subsequence (also denoted  $\{\bar{\mathbf{u}}^n(\cdot)\}$ ) converges  $L^1(0, T; L^1(\Omega))$ -weakly in  $L_w^\infty(0, T; L^\infty(\Omega))$  to a control  $\bar{\mathbf{u}}(\cdot)$  which is optimal for (2.2) in an interval  $0 \leq t \leq \bar{t}$ .

**Theorem 5.1.** *There exists  $(z_{on}, z_n) \in \mathbb{R} \times \Sigma(\Omega)$  such that*

$$(5.5) \quad z_{on} \leq 0, \quad z_n \in N_{Y_n}(\mathbf{y}_n(t_n, \zeta_n, \bar{\mathbf{u}}^n)), \quad |z_{on}| + \|z_n\| = 1$$

*( $N_{Y_n}(\mathbf{y}_n(t_n, \zeta_n, \bar{\mathbf{u}}^n))$  the normal cone to  $Y_n$  at  $\mathbf{y}_n(t_n, \zeta_n, \bar{\mathbf{u}}^n)$ ) such that if  $\mathbf{z}_n(t)$  is the solution of*

$$(5.6) \quad \begin{aligned} \mathbf{z}'_n(s) &= -\{A'_1 + S'_1(\varepsilon_n) \partial f(\mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n))^* S'_1(\varepsilon_n)\} \mathbf{z}_n(s) \\ &\quad - z_{on} S'_c(\varepsilon_n)^* \partial_y \mathbf{f}_0(S_c(\varepsilon_n) \mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n(s)), \quad \mathbf{z}_n(\bar{t}) = z_n \end{aligned}$$

and

$$(5.7) \quad \begin{aligned} H_n(\mathbf{y}(s, \zeta_n, \bar{\mathbf{u}}^n), z_{on}, \bar{\mathbf{z}}_n(s), \bar{\mathbf{u}}^n(s)) &= z_{on} \mathbf{f}_0(\mathbf{y}(s, \zeta_n, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n(s)) \\ &\quad + \langle (-A_1)^{1/2} \bar{\mathbf{z}}_n(s), (-A_c)^{1/2} \mathbf{y}(s, \zeta_n, \bar{\mathbf{u}}^n) \rangle \\ &\quad + \langle \bar{\mathbf{z}}_n(s), S_1(\varepsilon_n) \mathbf{f}(S_1(\varepsilon_n) \mathbf{y}(s, \zeta_n, \bar{\mathbf{u}}^n)) \rangle + \langle S_1(\varepsilon_n) \bar{\mathbf{z}}_n(s), \bar{\mathbf{u}}^n \rangle, \end{aligned}$$

then

$$(5.8) \quad H_n(\mathbf{y}(s, \zeta_n, \bar{\mathbf{u}}^n), z_{on}, \bar{\mathbf{z}}_n(s), \bar{\mathbf{u}}^n(s)) = \max_{v \in U} H_n(\mathbf{y}(s, \zeta_n, \bar{\mathbf{u}}^n), z_{on}, \mathbf{z}_n(s), v)$$

$$(5.9) \quad H_n(\mathbf{y}(s, \zeta_n, \bar{\mathbf{u}}^n), z_{on}, \bar{\mathbf{z}}_n(s), \bar{\mathbf{u}}^n(s)) = c_n$$

a.e. in  $0 \leq t \leq \bar{t}$ , with  $c_n = 0$  if the terminal time is free.

Theorem 5.1 is a particular case of [6, Theorem 6.3] except for (5.9); note that the target set contains interior points, which guarantees the third condition (5.5) (see the comments in [6] after Theorem 6.4).

We sketch a proof of (5.9). We note first that (3.1.n) can be written in the form  $\mathbf{y}'_n(t, \zeta_n, \bar{\mathbf{u}}^n) = A_c \mathbf{y}_n(t, \zeta_n, \bar{\mathbf{u}}^n) + \phi(t)$  where  $\mathbf{y}'_n(0, \zeta_n, \bar{\mathbf{u}}^n) = \zeta_n \in D(A_c)$  and where  $\phi(t) = S_c(\varepsilon_n) \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}_n(t, \zeta_n, \bar{\mathbf{u}}^n)) + S_\infty(\varepsilon_n) \mathbf{u}(t)$  and  $A_c \phi(t)$  are strongly measurable and bounded. It follows then from the theory of linear abstract parabolic equations that  $\mathbf{y}_n(t, \zeta_n, \bar{\mathbf{u}}^n) \in D(A_c)$  a.e. and is Lipschitz continuous and differentiable a.e. with  $\mathbf{y}'_n(t, \zeta_n, \bar{\mathbf{u}}^n) = A_c \mathbf{y}_n(t, \zeta_n, \bar{\mathbf{u}}^n) + S_c(\varepsilon_n) \mathbf{f}(S_c(\varepsilon_n) \mathbf{y}_n(t, \zeta_n, \bar{\mathbf{u}}^n)) + S_\infty(\varepsilon_n) \mathbf{u}(t)$ . Exactly the same considerations apply to  $\mathbf{z}_n(s)$  but, since the final condition  $\mathbf{z}_n$  is in  $\Sigma(\bar{\Omega})$ , only in intervals  $0 \leq s \leq \bar{t} - \delta$ ,  $\delta > 0$ . At any rate the function  $\langle \mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n), \mathbf{z}_n(s) \rangle$  is locally Lipschitz continuous in  $0 \leq s \leq t$  and its derivative can be directly calculated at any  $s$  where  $\mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n)$  and  $\mathbf{z}_n(s)$  can be differentiated; likewise, the derivative of  $\mathbf{f}_0(\mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n), \mathbf{u}(t))$  (as a function of  $s$ ) can be calculated a.e. Doing this and using the differential equations for  $\mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n)$  and  $\mathbf{z}_n(s)$ ,

$$(\partial/\partial s)|_{s=t} \phi(s, t) = (\partial/\partial s)|_{s=t} H_n(\mathbf{y}(s, \zeta_n, \bar{\mathbf{u}}^n), z_{0n}, \mathbf{z}_n(s), \mathbf{u}^n(t)) = 0$$

in a set  $c$  of full measure in  $[0, \bar{t}]$  independent of  $t$ . Since  $\phi(s, t)$  is  $s$ -Lipschitz continuous independently of  $t$  then the function  $\psi(s) = \inf_{0 \leq t \leq \bar{t}} \phi(s, t)$  is Lipschitz continuous, hence absolutely continuous, so that there exists a subset  $c_0 \subseteq c$ , still of full measure in  $[0, \bar{t}]$  where  $\psi'(s)$  exists. In this set we have

$$\begin{aligned} \frac{\psi(s+h) - \psi(s)}{h} &= \frac{\phi(s+h, s+h) - \phi(s, s)}{h} \\ &= \frac{\phi(s+h, s+h) - \phi(s+h, s)}{h} + \frac{\phi(s+h, s) - \phi(s, s)}{h}. \end{aligned}$$

If  $s \in c_0$  the left side and the second term on the right have a limit as  $h \rightarrow 0$ , thus so does the first term on the right. Since  $\phi(s+h, s+h) \leq \phi(s+h, s)$  for  $h$  of any sign, this limit must be zero. It follows that  $\psi'(s) = 0$  almost everywhere, thus  $\psi(s)$  is constant, proving (5.9). The statement for free terminal time problems involves using an auxiliary system and is essentially the same used in [3 pp. 130-32], so we omit the proof.

**Theorem 5.2.** *Let the target set  $Y$  be convex with nonempty interior. Then there exists  $(z_0, z) \in \mathbb{R} \times \Sigma(\bar{\Omega})$  such that*

$$(5.10) \quad z_0 \leq 0, \quad |z_0| + \|z\| = 1$$

and such that, if  $\mathbf{z}(s)$  solves

$$\begin{aligned} \mathbf{z}'(s) &= -\{A'_1 + \partial \mathbf{f}(\mathbf{y}(s, \zeta, \bar{\mathbf{u}}))^*\} \mathbf{z}(s) \\ &\quad - z_0 \partial_y \mathbf{f}_0(\mathbf{y}(s, \zeta, \bar{\mathbf{u}}), \bar{\mathbf{u}}(s)) \\ (5.11) \quad \mathbf{z}(\bar{t}) &= z \end{aligned}$$

and

$$(5.12) \quad \begin{aligned} H(\mathbf{y}(s, \zeta, \bar{\mathbf{u}}), z_0, \bar{\mathbf{z}}(s), \bar{\mathbf{u}}(s)) &= z_0 \mathbf{f}_0(\mathbf{y}(s, \zeta, \bar{\mathbf{u}}), \bar{\mathbf{u}}(s)) \\ &+ \langle (-A_1)^{1/2} \bar{\mathbf{z}}(s), (-A_c)^{1/2} \mathbf{y}(s, \zeta, \bar{\mathbf{u}}) \rangle \\ &+ \langle \bar{\mathbf{z}}(s), \mathbf{f}(\mathbf{y}(s, \zeta, \bar{\mathbf{u}})) \rangle + \langle \bar{\mathbf{z}}(s), \bar{\mathbf{u}} \rangle \end{aligned}$$

then

$$(5.13) \quad H(\mathbf{y}(s, \zeta, \bar{\mathbf{u}}), z_0, \bar{\mathbf{z}}(s), \bar{\mathbf{u}}(s)) = \max_{v \in I^+} H(\mathbf{y}(s, \zeta, \bar{\mathbf{u}}), z_0, \bar{\mathbf{z}}(s), v)$$

$$(5.14) \quad H(\mathbf{y}(s, \zeta, \bar{\mathbf{u}}), z_0, \bar{\mathbf{z}}(s), \bar{\mathbf{u}}(s)) = c$$

with  $c = 0$  if the problem is free terminal time.

Theorem 5.2 will be proved (of course) taking limits in Theorem 5.1. Note first that the sequences  $\{z_{on}\}$ ,  $\{z_n\}$  coming out of Theorem 5.1 are bounded; moreover, everything in  $H_n$  is bounded, (the second term by Corollary 3.4 and Corollary 4.2) so that  $c_n$  must be bounded. Hence, we may take a subsequence if necessary and assume that

$$(5.15) \quad c_n \rightarrow c, \quad z_{on} \rightarrow z_0, \quad z_n \rightarrow z,$$

the last  $\rightarrow$  understood  $C(\bar{\Omega})$ -weakly in  $\Sigma(\bar{\Omega})$ . To take limits directly in the Hamiltonian would not work, since  $\{\bar{\mathbf{u}}^n(\cdot)\}$  is just weakly convergent; we use the integrated version of (5.8) and (5.9). The first is

$$(5.16) \quad \begin{aligned} \int_c H_n(\mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n), z_{on}, \bar{\mathbf{z}}_n(s), \mathbf{v}(s)) ds \\ \leq \int_c H_n(\mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n), z_{on}, \bar{\mathbf{z}}_n(s), \bar{\mathbf{u}}^n(s)) ds \end{aligned}$$

for measurable sets  $c \subseteq [0, \bar{t}]$  and every  $\mathbf{v}(\cdot) \in C_{ad}(0, \bar{t}; U)$ ; the second

$$(5.17) \quad \int_c H_n(\mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}), z_0, \bar{\mathbf{z}}_n(s), \bar{\mathbf{u}}^n(s)) ds = c_n \text{meas}(c).$$

Clearly, we can go from (5.16) and (5.17) to (the integrated versions of) (5.13) and (5.14) if we can show that

$$(5.18) \quad \begin{aligned} \int_c H(\mathbf{y}(s, \zeta, \bar{\mathbf{u}}), z_0, \bar{\mathbf{z}}(s), \mathbf{v}(s)) ds \\ = \lim_{n \rightarrow \infty} \int_c H_n(\mathbf{y}(s, \zeta_n, \bar{\mathbf{u}}^n), z_{on}, \bar{\mathbf{z}}_n(s), \bar{\mathbf{v}}^n(s)) ds \end{aligned}$$

in just two cases:  $\mathbf{v}^n = \mathbf{v}(\cdot) \in C_{ad}(0, \bar{t}; U)$ , and  $\mathbf{v}^n(\cdot) = \bar{\mathbf{u}}^n(\cdot)$ . We concentrate on the second case. To take limits in the second term of the Hamiltonian (5.7) we use Corollary 3.4 for  $\mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n)$ , Corollary 4.2 for  $\bar{\mathbf{z}}^n(t)$  and the dominated convergence theorem. The second term is handled in the same way, using Corollary 3.4 and continuity of  $\mathbf{f}$ . Finally, the third term is disposed of using Corollary 4.2 (which

implies that  $\mathbf{z}_n(\sigma, \bar{\mathbf{u}}^n)$ , thus  $S_1(\varepsilon_n)\mathbf{z}_n(\sigma, \bar{\mathbf{u}}^n)$ , is convergent in  $L^1(0, \bar{t}; L^1(\Omega))$  and the  $L^1(0, \bar{t}; L^1(\Omega))$ -weak convergence of  $\bar{\mathbf{u}}^n(\cdot)$  in  $L_w^\infty(0, T; L^\infty(\Omega))$ . At this point, everything is over if  $z_{on} \rightarrow 0$ . Otherwise, we still have to prove that

$$(5.19) \quad \int_c \mathbf{f}_{on}(\mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n(s)) ds = \lim_{n \rightarrow \infty} \int_c \mathbf{f}_0(\mathbf{y}(s, \zeta, \bar{\mathbf{u}}), \bar{\mathbf{u}}(s)) ds.$$

This is obvious if we write

$$\mathbf{f}_{on}(\mathbf{y}(s, \zeta_n, \bar{\mathbf{u}}^n), \bar{\mathbf{u}}^n(s)) = \{c_n s - H_{n1}(\mathbf{y}_n(s, \zeta_n, \bar{\mathbf{u}}^n), z_{on}, \bar{\mathbf{z}}_n(s), \bar{\mathbf{u}}(s))\} / z_{on}$$

where  $H_{n1}$  comprises the terms in the Hamiltonian with whose convergence we can count on. This ends with (5.13) and (5.14). We must still prove the second condition (5.10) or, equivalently, that  $(z_0, z) \neq 0$ . If  $z_0 = \lim z_{on} \neq 0$  we are through; otherwise,  $\|z_n\| \rightarrow 1$ . We use the second condition (5.5), written in the form

$$\langle z_n, y \rangle \leq 0 \quad (y \in T_{Y_n}(\mathbf{y}_n(t_n, \zeta_n, \bar{\mathbf{u}}^n))).$$

$Y_n$  the tangent cone to  $Y$  at  $\mathbf{y}_n(t_n, \zeta_n, \bar{\mathbf{u}}^n)$ . Since  $Y$  is convex and has interior points all the  $T_{Y_n}(\mathbf{y}_n(t_n, \zeta_n, \bar{\mathbf{u}}^n))$  contain a common ball and  $z \neq 0$  by [6, Corollary 2.9], thus ending the proof of Theorem 5.2. We note that the convexity condition on  $Y$  is probably excessive, and it is likely that assumptions on the tangent cones of  $Y$  of the type used in [6, §2] suffice, although we have not been able to prove this.

## 6. Final comments

The Hilbert space setting in [3] is due to the use of "maximal regularity" results [1] for  $y'(t) = Ay(t) + f(t)$ . These have been extended to Banach spaces including  $L^p$  spaces in [2], but  $L^p$  extensions would still be very restrictive on nonlinearities. On the other hand, the results in [3] are not strictly less general than the ones here; for instance, the GEP is not needed.

Relaxed controls ((that is, measure-valued controls  $\mu(t)$ ) can be installed in any semilinear system  $y'(t) = Ay(t) + f(t, y(t), u(t))$  under very general assumptions [4]; with them, the system becomes  $y'(t) = Ay(t) + F(t, y(t))\mu(t)$ , linear in the control, and all the theory in this paper can be developed with minor changes: the "two-space" set up  $C(\bar{\Omega}), L^\infty(\Omega)$  and its dual  $L^1(\Omega), \Sigma(\bar{\Omega})$  is handled as in [6] or using the theory of Phillips adjoints (see [7]). However, it is essential that the semigroup  $S(t)$  generated by  $A$  be compact, as the arguments in this paper make clear. It seems possible, however, to extend at least some of the results to the case where the control  $u$  itself enters in the equation through a compact control term.

Another interesting line of generalization (requiring an analytic semigroup) would be to relax assumptions on the nonlinearity by means of "front-and-back smoothing with fractional powers", a trick well known for the Navier-Stokes equations; we assume here that  $((-A_1')^{-\beta})^* \mathbf{f}((-A_r)^{-\alpha}(\cdot))$ , not  $f(\cdot)$  is locally bounded

and locally Lipschitz continuous.  $D((-A_c)^\alpha)$ -valued solutions of the system (2.2) are constructed through the integral equation

$$\begin{aligned} (-A_c)^\alpha \mathbf{y}(t) = & (-A_c)^\alpha S_c(t)\zeta + \int_0^t (-A_c)^{\alpha+\beta} S_c(t-\tau) (-A_1')^{-\beta} \mathbf{f}(\mathbf{y}(\tau)) d\tau \\ & + \int_0^t (-A_c)^{-\alpha} S_\infty(t-\tau) \mathbf{u}(\tau) d\tau \end{aligned}$$

which can be solved in the customary way when  $\alpha + \beta < 1$ . The adjoint equation can be dealt with in the same way. However, convergence of  $(-A_1)^{1/2} \mathbf{z}_n(t, \bar{\mathbf{u}}^n)$  and of  $(-A_c)^{1/2} \mathbf{y}(t, \bar{\mathbf{u}}^n)$ , essential in the arguments in §4, seems out of reach unless  $\alpha + \beta < 1/2$ . This is excessive for dealing with realistic nonlinearities involving  $\text{grad } g(x)$ .

## References

1. L. DE SIMON, *Un'applicazione della teoria degli integrali singolari allo studio delle equazioni differenziali astratte del primo ordine*, Rend. Sem. Mat. Univ. Padova **34** (1964), 205–22.
2. G. DORE and A. VIENI, *On the closedness of the sum of two closed operators*, Math. Z. **196** (1987), 189–201.
3. H. O. FATTORINI, *Constancy of the Hamiltonian in infinite dimensional systems*, in Control and Estimation of Distributed Parameter Systems, (F. Kappel, K. Kunisch, W. Schappacher, Eds.), Int. Series Numerical Math., vol. 91, Birkhäuser, Basel, 1989, pp. 123–133.
4. H. O. FATTORINI, *Related controls in infinite dimensional control systems*, in Estimation and Control of Distributed Parameter Systems, (W. Desch, F. Kappel, K. Kunisch, Eds.), International Series of Numerical Mathematics, vol. 100, Birkhäuser, Basel, 1991, pp. 115–128.
5. H. O. FATTORINI, *Optimal control problems for distributed parameter systems governed by semilinear parabolic equations in  $L^1$  and  $L^\infty$  spaces*, in Optimal Control of Partial Differential Equations, (K.-H. Hoffmann, W. Krabs, Eds.), Springer Lecture Notes in Control and Information Sciences, vol. 149, 1991, pp. 68–80.
6. H. O. FATTORINI, *Optimal control problems in Banach spaces*, Appl. Math. Optim. **28** (1993), 225–257.
7. H. O. FATTORINI, *Relaxation in semilinear infinite dimensional control systems*, in L. Markus Festschrift, (K. D. Elworthy, W. N. Everitt, E. B. Lee, Eds.), Marcel Dekker, New York, 1993, pp. 505–522.
8. H. O. FATTORINI and H. FRANKOWSKA, *Necessary conditions for infinite dimensional control problems*, Math. Control, Signals and Systems **4** (1991), 41–67.
9. H. O. FATTORINI and T. MURPHY, *Optimal problems for nonlinear parabolic boundary control systems*, to appear in SIAM J. Control & Optimization.
10. H. FRANKOWSKA, *Some inverse mapping theorems*, Ann. Inst. Henri Poincaré **7** (1990), 183–234.
11. D. HENRY, "Geometric Theory of Semilinear Parabolic Equations", Springer, Berlin 1981.
12. E. HILLE and R. S. PHILLIPS, "Functional Analysis and Semi-Groups", AMS Colloquium Publ., vol. 31, Providence, 1957.

H. O. Fattorini  
Department of Mathematics  
University of California  
Los Angeles, California 90024, USA

## RATE DISTRIBUTION MODELING FOR STRUCTURED HETEROGENEOUS POPULATIONS

BEN G. FITZPATRICK

Center for Research in Scientific Computation  
Department of Mathematics  
North Carolina State University

**ABSTRACT.** We propose a modeling strategy for structured populations, in which individuals are not necessarily identical. The heterogeneity is obtained by modeling the population as comprising homogeneous subpopulations. By using a vector measure, we combine the subpopulations with an abstract integral to obtain the density of the population. We show that this approach leads to a semigroup formulation of the dynamics in a space of vector measures, and we develop some estimation methods for determining the initial structure from observed data.

1991 *Mathematics Subject Classification.* 46G10, 92D25, 92D40

*Key words and phrases.* Rate distributions, vector measures, semigroups, estimation.

### 1. Introduction

Many biological applications of dynamical systems and control theory involve populations that are structured in terms of size, age, or spatial distribution. From predicting future populations to developing optimal harvesting strategies, models provide important information for biologist. Thus, it is crucial to develop models that predict population behavior in an accurate manner. This paper focuses on a particular aspect of model improvement, that of incorporating individual-based information into aggregate, population-wide models.

The original motivation for this work involved certain behavior exhibited by observed data for a size structured population of mosquitofish, behavior which is inconsistent with the commonly used Sinko-Streifer model:

$$v_t + (gv)_x = -\mu v, \quad x_0 < x < x_1, t > 0.$$

Here  $v = v(t, x)$  denotes the density of individuals of size  $x$  at time  $t$ ,  $g = g(t, x)$  denotes individual growth rate, and  $\mu = \mu(t, x)$  is the mortality rate. The observed data, as discussed in [BBKW, BF], exhibits a dispersion in size as time increases,

---

Research supported in part under AFOSR contract F49620-93-1-0153.



and in some cases unimodal initial densities develop into bimodal densities at later time: the Sinko-Streifer model does not predict these phenomena without biologically unrealistic assumptions on the parameters.

In order to improve model predictions, various researchers (see [BBKW, BF]) have attempted to examine the model on the individual level. In general, individuals do not have identical growth rates (a basic assumption for the Sinko-Streifer model). To introduce individual variations, we model the population as being composed of homogeneous subpopulations, with each subpopulation obeying a Sinko-Streifer law having different parameters. The approach in [BBKW, BF] involves subpopulation densities  $v(t, x; g)$ , with the growth rate parameter  $g$  differentiating the subpopulations. The subpopulations are then combined by integrating the densities with respect to a measure on the space of parameters:

$$u(t, x) = \int_G v(t, x; g) dP(g).$$

The measure  $P$ , called the growth rate distribution, represents the proportion of individuals having a given growth rate.

In this paper, we consider the more general problem of a population whose dynamics are modeled with a parameter-dependent  $C_0$  semigroup  $T(t; q)$  on a Hilbert space  $X$ . These operators typically arise as solution operators for a differential equation such as the Sinko-Streifer equation given above. This general approach allows us to consider several different kinds of models from many differential equations - age, size, spatial structure models, or a combination. Within the context of population models, the parameter  $q$  typically comprises several individual rate parameters, such as growth, mortality, and fecundity. The rate distribution idea then is to model the population as a combination of subpopulations which are modeled with the original semigroup dynamics  $T(t; q)$ . The question then becomes how to build the whole population from the subpopulations.

Suppose for the moment that the number of subpopulations is finite. Then, we denote the densities by  $v(t, x; q_i) = T(t, q_i)\varphi_i(x)$ , for  $1 \leq i \leq n$ . Then the density  $u(t, x)$  of the whole population is given by

$$(1.1) \quad u(t, x) = \sum_{i=1}^n T(t, q_i)\varphi_i(x).$$

In [BF], the initial densities were assumed to satisfy

$$(1.2) \quad \varphi_i = p_i \varphi,$$

for some  $\varphi \in X$ , and some collection  $\{p_i\}$  with  $p_i \geq 0$ , and  $\sum p_i = 1$ . Then the population density can be given in the form

$$u(t, x) = \sum_{i=1}^n T(t, q_i)\varphi(x)p_i = \int_Q T(t; q)\varphi(x) dP(q),$$

where  $P$  is the discrete probability measure on  $Q$  with support  $\{q_i\}$  and weights  $\{p_i\}$ . The integral form is generalized naturally to the case mentioned above in which  $P$  is any probability measure on  $Q$ , and the estimation theory of [BF] generalizes in a straightforward manner for this problem.

To generalize this model further, we relax (1.2), so that the subpopulations may have different initial structure as well as distinct rate parameters. Thus we are led to a generalization of the form

$$(1.3) \quad u(t) = \int_Q T(t; q) \varphi(q) dP(q),$$

where the  $X$  valued function  $\varphi(q)$  gives the parameter dependent initial structure. A further step is to write

$$(1.4) \quad u(t) = \int_Q T(t; q) d\mathbf{m}(q),$$

where  $\mathbf{m}$  is a vector-valued measure on  $Q$  (taking values in  $X$ ). Intuitively, we have  $\mathbf{m}(dq) = \varphi(q) P(dq)$ . Also, when  $X = L^2$ , say,  $\int_a^b \mathbf{m}(A)(x) dx$  denotes the number of individuals whose structure variable  $x$  is between  $a$  and  $b$ , and whose parameter  $q$  lies in the set  $A \subset Q$ . Note that formula (1.4) above involves integrating an operator valued function with respect to a vector measure. There are some subtle questions of measurability in both (1.3) and (1.4) that must be resolved from properties of the original semigroup model (see [F]).

The advantage to using the form (1.4) over (1.3) is not in a greater level of generality, for under rather general conditions a vector measure can be expressed in terms of a vector valued density and a scalar measure. However, as we shall see below, the vector measure approach leads directly to a semigroup formulation of the dynamics in a vector measure space, and hence we may view the state as possessing not only the original (size, age, or spatial) structure but also the individual (rate) structure.

We shall consider the population state at time  $t$  to be  $\mathbf{m}_t$ , which can be given in mild form at

$$(1.5) \quad \mathbf{m}_t = T(t)\mathbf{m}_0 + \int_0^t T(t-s)\mathbf{f}(s) ds,$$

where  $\mathbf{f}$  is a "forcing function" modeling changes in parameter structure due to external environmental changes. Below we shall derive the semigroup  $T$  from  $T(t; q)$ : the idea is that  $T(t)\mathbf{m}_0(dq) = T(t; q)\varphi(q)\mu(dq)$ , where we have  $\mathbf{m}_0(dq) = \varphi(q)\mu(dq)$ .

The inverse problem is to estimate the measure  $\mathbf{m}_0$  and the forcing function  $\mathbf{f}$  from observations of the population. In size structured populations, one typically obtains "histogram" data which represents numbers of individuals whose size lies in given "size bins." We assume that the observations are of the form  $\{\hat{u}(t_i): 1 \leq$

$i \leq n\} \subset Z$ , where the data space  $Z$  is a Hilbert space, and we use the least squares criterion

$$J(\mathbf{m}_0, \mathbf{f}) = \sum_{k=1}^n \|\hat{u}(t_k) - \mathcal{C}\mathbf{m}_t(Q)\|_Z^2,$$

where  $\|\cdot\|_Z$  denotes the norm in  $Z$ , and where  $\mathcal{C}: X \rightarrow Z$  is the observation mapping (e.g., a projection onto the step functions in the histogram case). Computation of solutions to this inverse problem requires several approximations — for the semigroup, the vector integral in (1.5), and the vector measure itself. It does, however, provide an advantage over identification of the parameter  $q$ : the least squares problem for the measure is a linear least squares problem. A very similar idea, relaxed control, has been used in control theory for quite some time (see, e.g. [W]).

A computational advantage in this approach is that the problem contains a high level of parallelism. Different subpopulations can be simulated independently; hence, parallel computing platforms can be used very efficiently.

This paper is organized as follows. In Section 2, we discuss the functional analytic tools necessary to put (1.5) on a rigorous foundation. Section 3 contains the approximation methods for inverse problems, and some remarks on future work are contained in Section 4.

## 2. Semigroup Formulation of the Model

Our goal in this section is to extend the dynamics of a population comprising identical individuals to a heterogeneous population, distributed not only in its original structure but also in its rate structure. We begin with the original dynamics,

$$(2.1) \quad \dot{v}(t; q) = A(q)v(t; q),$$

where  $A$  is the infinitesimal generator of a  $C_0$  semigroup  $T(t; q)$  on a Hilbert space  $X$ . In the extension we are seeking, the solutions  $v(t; q)$  represent the subpopulation density having parameter  $q$ . Of course, when we refer to this function of  $q$  as a density, we need a measure with respect to which  $v$  is a density. We also need an extended state space which incorporates the additional structure.

One natural approach to a state space formulation is to take the initial population to be an  $X$ -valued measure. The space  $\mathcal{M}$  of regular, countably additive  $X$ -valued Borel measures of bounded variation on  $Q$  forms a Banach space which can be identified with  $C(Q, X)^*$ . We include here some definitions and notation which will be very useful below. The interested reader may consult [DU] or [D] for the details of vector measures. First, we recall that a *countably additive vector measure* (or more concisely, a *vector measure*) is a function from a  $\sigma$ -algebra  $\mathcal{F}$  of sets in  $Q$  to  $X$  satisfying

$$\mathbf{m}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbf{m}(E_i)$$

for every collection  $\{E_i\} \subset \mathcal{F}$  of disjoint sets. The limit in the infinite sum holds in the  $X$  norm sense. We define the *variation of  $\mathbf{m}$* ,  $|\mathbf{m}|$ , by

$$|\mathbf{m}|(E) = \sup_{\pi \in \Pi} \sum_{E_i \in \pi} \|\mathbf{m}(E_i)\|_X,$$

where  $\Pi$  is the collection of finite partitions of  $E$ . If  $|\mathbf{m}|(Q) < \infty$ , we say that  $\mathbf{m}$  is of *bounded variation*. The *semivariation* is defined by

$$\|\mathbf{m}\|(E) = \sup \left\{ |\langle x, \mathbf{m}(E) \rangle| : x \in X, \|x\|_X = 1 \right\},$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product in  $X$ . We say that  $\mathbf{m}$  is of *bounded semivariation* if  $\|\mathbf{m}\|(Q) < \infty$ . A vector measure is of bounded semivariation if and only if its range is bounded in  $X$  (p. 5 of [DU]) so that we refer to measures of bounded semivariation simply as *bounded measures*. Unless otherwise stated, integrals of vector valued functions are taken to be Bochner integrals (see [DU, Chapter 2]).

Measures of bounded variation can also be expressed as  $\mathbf{m}(dq) = \varphi(q) \mu(dq)$ , where  $\varphi \in L^1(Q, \mu; X)$  and  $\mu$  is a positive real-valued measure on  $Q$ , such that  $0 \leq \mu(A) \leq \|\mathbf{m}\|(A)$ , for each measurable  $A$ . This fact follows from a theorem of Bartle, Dunford, and Schwarz and the fact that Hilbert spaces have the Radon-Nikodym property (see [DU, pp. 14, 61, and 100]).

Another state space possibility is given by  $C(Q; X) \times M$ , where  $M (= C(Q)^*)$  denotes the finite real-valued Borel measures on  $Q$ , for  $Q$  a compact separable metric space. This approach has been successfully employed in [BKW] within the context of model development.

Here we shall focus on the vector measure approach, which has proven effective in inverse problems (see [F] and the following sections). We begin by extending the original semigroup to the vector measure space. The following result is the first necessary step.

**Lemma 2.1.** *Suppose that  $X$  is a Hilbert space and that  $T(t; q)$  is a semigroup on  $X$  that is strongly jointly continuous on  $[0, \infty) \times Q$ , where  $Q$  is a compact, separable metric space. Suppose that  $\mu$  is a finite regular nonnegative measure on (the Borel sets of)  $Q$  and that  $\varphi \in L^1(Q, \mu; X)$ . Then, for each  $t$ , the map  $q \mapsto T(t, q)\varphi(q)$  is in  $L^1(Q, \mu; X)$ , and the vector measure  $\mathbf{n}$  given by  $\mathbf{n}(A) = \int_A T(t, q)\varphi(q) \mu(dq)$  is in  $\mathcal{M}$ .*

**Proof.** First, we note that if the above map is measurable, then we have that

$$\begin{aligned} \int_Q \|T(t; q)\varphi(q)\|_X \mu(dq) &\leq \int_Q \|T(t; q)\| \|\varphi(q)\|_X \mu(dq) \\ &\leq M e^{\omega t} \int_Q \|\varphi(q)\|_X \mu(dq) < \infty, \end{aligned}$$

since  $T$  is jointly strongly continuous (see [DU, F]). Thus, to prove that the above map is  $L^1$ , it remains to show measurability.

We recall that an  $X$ -valued function is measurable if and only if it is the ( $\mu$ -a.e.) pointwise strong limit of simple functions (see [DU]). Given that  $\varphi \in L^1(Q, \mu; X)$ , we may choose a sequence of simple functions  $\varphi^n \rightarrow \varphi$  strongly, for  $\mu$ -a.e.  $q \in Q$ . Next, for each  $n$ , we choose a finite collection of balls of radius  $1/n$ ,  $B(q_i^n, 1/n)$ , for  $1 \leq i \leq K_n$  that cover  $Q$ . We redefine these balls in order to make them disjoint, by setting  $B_1^n = B(q_1^n, 1/n)$ , and letting

$$B_k^n = B(q_k^n, 1/n) \cap (B_1^n)^c \cap \dots (B_{k-1}^n)^c.$$

We then set

$$(2.2) \quad \psi^n(q) = \sum_{k=1}^{K_n} \sum_{i=1}^{N_n} T(t, q_k^n) \phi_i^n \chi_{A_i^n \cap B_k^n},$$

where  $\varphi = \sum_{i=1}^{N_n} \phi_i^n \chi_{A_i^n}$ . Clearly,  $\psi^n$  is a simple function. Moreover, if  $q_0$  is a point for which  $\varphi^n(q_0) \rightarrow \varphi(q_0)$ , then there is a sequence of sets  $A_{i_n}^n \cap B_{k_n}^n$  that contain  $q_0$  and whose radii are tending to 0. By the strong continuity of  $T$ , together with the convergence of  $\varphi^n$ , we have that  $T(t; q_{k_n}^n) \varphi^n(q_0) \rightarrow T(t; q_0) \varphi(q_0)$ , as desired.

For the last claim, we have that  $\mathbf{n}(A) = \int_A T(t, q) \varphi(q) \mu(dq)$  is a countably additive, bounded variation vector measure (from [DU, p. 46]) and that in fact

$$|\mathbf{n}|(A) = \int_A \|T(t, q) \varphi(q)\|_X \mu(dq).$$

To prove regularity of  $\mathbf{n}$ , we must show that for every measurable set  $A$  and for every  $\varepsilon > 0$ , there exist sets  $K$ , compact, and  $O$ , open satisfying  $K \subset A \subset O$ , and  $|\mathbf{n}|(O - K) < \varepsilon$ .

Note that if  $\varepsilon > 0$ , since  $q \mapsto T(t, q) \varphi(q)$  is in  $L^1(Q, \mu; X)$ , there exists  $\delta > 0$  such that if  $B$  is measurable and  $\mu(B) < \delta$ , then  $|\mathbf{n}|(B) < \varepsilon$ . The regularity of  $\mu$  then provides the result.

Using this lemma, we define the operators  $\mathcal{T}(t)$  on  $\mathcal{M}$  by setting

$$(2.3) \quad \mathcal{T}(t) \mathbf{m}(dq) = T(t; q) \varphi(q) \mu(dq).$$

To prove that this definition gives us a semigroup on  $\mathcal{M}$ , we must first show that  $\mathcal{T}(t)$  is well defined. That is, suppose  $d\mathbf{m} = \varphi d\mu = \psi d\nu$ . (In fact, many such distinct representations exist; see [DU, p. 269, Corollary 3]). We must show that both representations lead to the same linear operator; i.e.,

$$T(t; q) \varphi(q) \mu(dq) = T(t; q) \psi(q) \nu(dq).$$

Toward that end, we take  $B_k^n$  and  $q_k^N$  as in the above proof, and we define a sequence of operators

$$T^n(t; q)x = \sum_{k=1}^{K_n} \chi_{B_k^n}(q) T(t; q_k^N)x,$$

which converges strongly to  $T(t; q)$ . In a manner similar to (2.2), we set

$$\mathbf{n}_1^n(A) = \sum_{k=1}^{K_n} \int_{A \cap B_k^n} T(t; q_k^N) \varphi \mu(dq) = \sum_{k=1}^{K_n} T(t; q_k^N) \int_{A \cap B_k^n} \varphi \mu(dq),$$

and

$$\mathbf{n}_2^n(A) = \sum_{k=1}^{K_n} \int_{A \cap B_k^n} T(t; q_k^N) \psi \nu(dq) = \sum_{k=1}^{K_n} T(t; q_k^N) \int_{A \cap B_k^n} \psi \nu(dq).$$

From the hypothesis on  $\varphi, \psi, \mu$ , and  $\nu$ , we see that  $\mathbf{n}_1^n = \mathbf{n}_2^n$ . Moreover, by the dominated convergence theorem, we have that  $\mathbf{n}_i^n(A) \rightarrow \mathbf{n}_i(A)$  for  $i = 1, 2$ , where

$$\mathbf{n}_1(A) = \int_A T(t, q) \varphi(q) \mu(dq)$$

and  $\mathbf{n}_2$  is similarly defined. Since  $\mathbf{n}_1^n = \mathbf{n}_2^n$ , we must have  $\mathbf{n}_1 = \mathbf{n}_2$ , which implies that  $T(t)$  is well defined.

We also note that since

$$|\mathbf{n}|(A) = \int_A \|T(t, q) \varphi(q)\|_X \mu(dq) \leq M e^{\omega t} |\mathbf{m}|(A),$$

$T(t)$  is a bounded linear operator on  $\mathcal{M}$ . Using these facts we have the following.

**Theorem 2.2.** *Under the hypotheses of Lemma 2.1, we have that  $T(t), t \geq 0$  is a strongly continuous semigroup of linear operators on  $\mathcal{M}$ .*

**Proof.** The semigroup property is easily seen to follow from (2.3), as is the fact that  $T(0) = I$ . We must now argue strong continuity. Note that

$$|T(t)\mathbf{m} - \mathbf{m}|(Q) = \int_Q \|T(t, q) \varphi(q) - \varphi(q)\|_X \mu(dq),$$

which goes to 0 as  $t$  goes to 0, by the dominated convergence theorem. Thus strong continuity is proved.

From this theorem we obtain a model for dynamics in the vector measure state space,  $\mathcal{M}$ , through the mild form given in (1.5):

$$\mathbf{m}_t = T(t)\mathbf{m}_0 + \int_0^t T(t-s)\mathbf{f}(s) ds,$$

where  $\mathbf{f} \in L^1([0, t_f]; \mathcal{M})$ . The state vector now is  $\mathbf{m}_t$  which contains structure in both the parameter space and the original space  $X$ . Moreover, the mild form also contains a forcing function which can be used to model externally induced parameter changes. Having developed a semigroup formulation for the extended dynamics, we now turn to inverse problems.

### 3. Inverse Problems for the Rate Distribution Model

It is typically the case in a population biology application that the individual rate constants cannot be measured directly. Thus, in population wide observations, one would be unlikely to have access to measurements of parameter variation in the population. On the other hand, size and spatial structure are observable in many cases. Thus, we are confronted with an inverse problem in which the observations  $\hat{u}(t_i) \in Z$  correspond to  $\mathcal{C}(\int_Q T(t_i; q)\varphi(q)\mu(dq))$ , where  $\mathcal{C}: X \rightarrow Z$  is an observation operator and the data space  $Z$  is a separable Hilbert space. In many population biology examples, the data are histograms giving the number of individuals whose size (age, position) lies in particular intervals. In such a case  $Z$  is finite dimensional and  $\mathcal{C}$  is an integral "averaging" operator. The integral in  $q$  denotes the summing of the population over the parameter space: we observe the total population, structured only in  $X$ . We thus pose the inverse problem as the determination of  $\mathbf{m}$  from such measurements. We shall use the least squares approach, in which we seek to minimize

$$(3.1) \quad J(\mathbf{m}) = \sum_{i=1}^n \|\mathcal{C}T(t_i)\mathbf{m}(Q) - \hat{u}(t_i)\|_Z,$$

which, we note, is an infinite dimensional linear least squares problem.

A possible approach to the minimization is through standard linear least squares theory and the normal equations  $B^*B\mathbf{m} = \hat{u}$ , with  $B = (T(t_1), \dots, T(t_n))^T \in \mathcal{L}(X^n)$ , and  $\hat{u} = (\hat{u}(t_1), \dots, \hat{u}(t_n))^T \in X^n$ . We do not have at this point any general conditions on the original semigroup under which the normal equations have a unique solution. Hence, we take here a different approach.

We treat the problem as a constrained minimization of  $J$  over a set  $\mathcal{M}_{ad}$  of admissible parameters. Since the unit ball in  $C(Q, X)^*$  is weak-\* compact (by Alaoglu's theorem), we take  $\mathcal{M}_{ad} \subset \mathcal{M}$  to be the measures with variation bounded by  $M$ , which is a fixed positive number. Under the assumption that the adjoint semigroup,  $T^*(t; q)$ , is strongly continuous in  $t$  and  $q$ , we have that  $\mathbf{m}^N \rightharpoonup \mathbf{m}$  weak-\* implies that  $T(t)\mathbf{m}^N(Q) \rightarrow T(t)\mathbf{m}(Q)$  weakly in  $X$ , for

$$\begin{aligned} \langle x, \int_Q T(t; q)\varphi(q)\mu(dq) \rangle &= \int_Q \langle x, T(t; q)\varphi(q) \rangle \mu(dq) \\ &= \int_Q \langle T^*(t; q)x, \varphi(q) \rangle \mu(dq). \end{aligned}$$

Since  $T^*$  is strongly continuous,  $q \mapsto T^*(t, q)x \in C(Q, X)$ , and the weak convergence becomes apparent. From this convergence we obtain the following.

**Theorem 3.1.** *Suppose that  $\mathcal{C} \in \mathcal{L}(X, Z)$  is compact. Under the assumptions of Theorem 2.2, together with strong continuity of the adjoint semigroup  $T^*(t; q)$ , the functional  $J$  is continuous with respect to the weak-\* topology on  $\mathcal{M}_{ad}$ ; hence,  $J$  attains a minimum over  $\mathcal{M}_{ad}$ .*

**Proof.** First, note that the weak-\* topology on  $\mathcal{M}_{ad}$  and the weak topology of  $\mathcal{T}(t)(\mathcal{M}_{ad}) \subset X$  are metrizable (see, e.g., [R, p. 203]). Thus, the above argument shows that  $\mathbf{m} \mapsto \mathcal{T}(t)\mathbf{m}$  is continuous with respect to these topologies. Since  $\mathcal{C}$  is compact, it maps weakly convergent sequences to strongly convergent sequences, and we have the desired result.

The optimization problem posed is an infinite dimensional one, so we need some approximation methods for computational purposes. We assume that we have for the original problem a numerical scheme that produces a sequence of semigroups  $T^N(t; q)$  defined on finite dimensional subspaces  $X^N$  of  $X$ , taken to be of the form  $X^N = \text{span}\{\psi_1^N, \dots, \psi_N^N\}$ , with  $P^N$  being the orthogonal projection from  $X$  to  $X^N$ . From these semigroups we construct cost functionals

$$J^N(\mathbf{m}) = \sum_{k=1}^n \|\hat{u}(t_k) - \mathcal{C}T^N(t)\mathbf{m}(Q)\|_Z^2,$$

where  $T^N$  is defined in the obvious manner. This cost functional must be minimized over a finite dimensional set. Let  $\{q^1, q^2, \dots\}$  be a countable, dense subset of  $Q$ . We set

$$\mathcal{M}^N = \left\{ \sum_{j,k=1}^N a_{jk} \psi_j^N \delta_{q^k}(dq) : \left| \sum_{j,k=1}^N a_{jk} \psi_j^N \delta_{q^k} \right|(Q) \leq M \right\},$$

which is a finite dimensional subset of  $\mathcal{M}_{ad}$ . Furthermore (as in [F]), any element of  $\mathcal{M}_{ad}$  can be approximated in the weak-\* sense by a sequence from these sets. With these observations, we may obtain the following approximation result.

**Theorem 3.2.** *Assume that the hypotheses of Theorem 3.2 hold. Furthermore let the sequence  $T^N(t; q)$  satisfy  $T^N(t; q)P^N x \rightarrow T(t; q)x$ , and  $(T^N)^*(t; q)P^N x \rightarrow T^*(t; q)x$  strongly in  $X$  and uniformly on  $Q \times [0, \tau]$ . Then minimizers of  $J^N$  over  $\mathcal{M}^N$  converge subsequentially in the weak-\* topology of  $\mathcal{M}_{ad}$  to minimizers of  $J$ .*

*Proof.* In view of the abstract least squares theory of Banks (see, e.g., [BK, pp. 144-5]) it only remains to show that  $J^N(\mathbf{m}^N) \rightarrow J(\mathbf{m})$ , whenever  $\mathbf{m}^N \rightarrow \mathbf{m}$  in the weak-\* sense. Toward that end, we note that

$$\begin{aligned} \langle x, \int_Q T^N(t; q) P^N \varphi^N(q) \mu^N(dq) \rangle &= \int_Q \langle x, T^N(t; q) P^N \varphi^N(q) \rangle \mu^N(dq) \\ &= \int_Q \langle P^N x, T^N(t; q) P^N \varphi^N(q) \rangle \mu^N(dq) \\ &\quad - \int_Q \langle x, T(t; q) \varphi^N(q) \rangle \mu^N(dq) \\ &\quad + \int_Q \langle x, T(t; q) \varphi^N(q) \rangle \mu^N(dq) \end{aligned}$$



$$\begin{aligned}
&= \int_Q \langle (T^N(t; q))^* P^N x, \varphi^N(q) \rangle \mu^N(dq) \\
&\quad - \int_Q \langle T^*(t; q)x, \varphi^N(q) \rangle \mu^N(dq) \\
&\quad + \langle x, \int_Q T(t; q) \varphi^N(q) \mu^N(dq) \rangle \\
&\rightarrow \langle x, \int_Q T(t; q) \varphi(q) \mu(dq) \rangle.
\end{aligned}$$

which implies  $\mathcal{T}^N(t)\mathbf{m}^N \rightarrow \mathcal{T}(t)\mathbf{m}$  weakly in  $X$ . As in Theorem 3.1, we then obtain the desired convergence for  $J^N$ .

#### 4. Conclusions

We have examined here a semigroup formulation of a rate distribution model for structured populations. Being based merely on an original semigroup formulation for subpopulations of identical individuals, this model is quite general. Using basic properties of vector measures, we studied the basic well-posed question for the semigroup formulation of the distributed model, and developed an estimation framework for fitting the model to data.

We have performed a variety of computational examples based on the Sinko-Streifer problem mentioned above. The implementation involves solving the Sinko-Streifer PDE for many growth rate functions, and we have successfully used the 128-processor Intel iPSC/860 at Oak Ridge National Laboratory for the computations. These results, reported in [BFZ], demonstrate not only the computational feasibility of the techniques we have described, but also the capability of obtaining high quality fits to real observed size structured data that were not previously treatable with inverse techniques.

Future studies will focus on vector measure formulations for nonlinear problems (such as [H]) as well as models which incorporate subpopulation mixing (as in [BKW]).

#### References

- [BBKW] Banks, H. T., L. Botsford, F. Kappel, and C. Wang. Modeling and Estimation in Size Structured Population Models, *Proceedings 2<sup>nd</sup> Course on Mathematical Ecology*, (Trieste, December 8-12, 1986), World Press, Singapore (1988), 521-541.
- [BF] Banks, H. T., and B. G. Fitzpatrick. Estimation of Growth Rate Distributions in Size Structured Population Models, *Q. Applied Math.*, **49**, no 2, 1991, pp. 215-235.
- [BFZ] Banks, H. T., B. G. Fitzpatrick, and Y. X. Zhang. "A Parallel Algorithm for Rate Distribution Estimation in Size Structured Population Models," to appear.
- [BKW] Banks, H. T., F. Kappel, and C. Wang. "Semigroup formulation of a heterogeneous population model and its nonlinear perturbations," to appear.
- [BK] Banks, H. T., and K. Kunisch. "Estimation Techniques for Distributed Parameter Systems," Birkhäuser, Boston, 1989.

- [DU] Diestel, J., and J. J. Uhl, Jr. "Vector Measures," Mathematical Surveys **15**, American Mathematical Society, Providence, 1977.
- [D] Dinculeanu, N. "Vector Measures," Pergamon Press, Oxford, 1967.
- [DS] Dunford, N., and J. T. Schwartz. "Linear Operators, Part I: General Theory." Wiley Interscience, New York, 1966.
- [F] Fitzpatrick, B. G. Modeling and Estimation Problems for Structured Heterogeneous Populations, *J. Math. Anal. Appl.*, **172** no. 1, 1993, pp. 73-91.
- [H] Hoyer, W. "A Size Structured Population Model with Dispersion," *J. Math. Anal. Appl.*, to appear.
- [W] Warga, J. "Optimal Control of Differential and Functional Equations," Academic Press, New York, 1972.

Ben G. Fitzpatrick  
Center for Research in Scientific Computation  
Department of Mathematics  
North Carolina State University  
Raleigh, NC 27695-8205, USA

## A MODEL FOR A TWO-LAYERED PLATE WITH INTERFACIAL SLIP

SCOTT W. HANSEN

Department of Mathematics  
Iowa State University

**ABSTRACT.** In this paper we derive a model for a two-layered plate in which slip can occur at the interface. We assume that a "glue" layer of negligible thickness bonds the two adjoining surfaces in such a way that the restoring force created by the glue is proportional to the amount of slippage. Within each plate the assumptions of Timoshenko beam theory (namely, that the straight filaments orthogonal to each center sheet at equilibrium remain straight during deformation) are applied and the equations of motion are derived through the principle of virtual work. We relate the resulting system to the Mindlin-Timoshenko-Reissner plate system and also to the Kirchhoff plate system by singular perturbations involving passing the shear stiffness parameter and the glue strength parameter to infinity.

1991 *Mathematics Subject Classification.* 73K10, 73K20

*Key words and phrases.* Multi-layer plate, Mindlin plate, Reissner plate.

### 1. Introduction

Over the past decades, composite materials have found increasingly broad application in many areas from design of sporting goods to space structures. These materials can be designed to produce either favorable damping characteristics or high strength to mass ratios, and often for a combination of each. Consequently there has been a considerable amount of effort in modelling these structures and analyzing the dynamical properties thereof.

Models for multilayered beams and plates have existed in the literature since at least the late fifties [Yu], [Ke]. Since then, numerous more elaborate models have been developed (see [He], [Di], [Me], [YD] and the numerous references therein). For those models which were constructed with the idea of modelling damping the usual approach has been to assume the material consists of three separate layers which are bonded together (no-slip at the interfaces). The outer layers are modelled by plates which allow little or no shear while the inner layer is modelled as a material with a "complex shear modulus" but usually without other material assumptions. Kinematic restrictions are then artificially imposed which couple the

strains or stresses in each layer with one another and a plate model involving only the transverse displacement is produced.

In this paper we derive a model for a two-layered plate in which slip can occur at the interface. We imagine that a "glue" layer of negligible thickness bonds the two adjoining surfaces in such a way that the restoring force created by the glue is proportional to the amount of slippage. Within each plate the assumptions of Timoshenko beam theory [Ti] (namely, that the straight filaments orthogonal to each center sheet at equilibrium remain straight during deformation) are applied and the equations of motion are derived through the principle of virtual work.

In the approach used here, we impose no *a-priori* kinematic restrictions coupling the motion of the upper plate to the lower one. (Although we do assume within each plate standard assumptions of Timoshenko beam theory). However, at least for the present article we will assume a geometric symmetry in that we consider only the case where the upper and lower plates are identical. This allows for a decoupling of the in-plane motions from the bending motions.

Initially we obtain a system of nine equations which describe the dynamics of the composite plate. The part describing the bending has five equations in the transverse displacement, the *effective* rotation angles and two for the components of the slip. We show that this system is an analog of the Mindlin-Timoshenko-Reissner plate system (called the Mindlin-Timoshenko plate in [LL], [La]) in the sense that 1) when there is no glue our system reduces to an MTR system, 2) when the strength of the glue tends to infinity, solutions of our system converge in a weak sense to solutions of an MTR system.

By letting the shear-stiffness parameter tend to infinity we obtain a limiting system of three equations which is analogous to the Kirchhoff system in the same sense that our original system is analogous to the MTR system.

A major reason for including slip (as opposed to including a center layer for the glue) is that one can then easily introduce frictional forces into the dynamics of the slip. If one models this force as simple sliding friction (frictional force proportional to velocity), one obtains two new *structurally damped* plate models. It is known that much of the dissipation in composite materials is in fact due to friction between adjacent fibers. Thus a goal of this research is to characterize the damping one obtains in plates due to internal friction. However in the present paper we concern ourselves with the case in which no frictional forces are included.

## 2. Basic assumptions

Much of our development and notation will follow that of Lagnese and Lions [LL, chapters 1,2].

Our plate consists of two identical nonhomogeneous (homogeneous in the transverse direction), isotropic thin plates which are in contact with each other on a *middle surface* which occupies the region  $\Omega \subset \mathbb{R}^2$  at equilibrium. Each plate is assumed to have uniform thickness  $h/2$ .

Let us use the rectangular coordinates  $\underline{x} = (x_1, x_2)$  to denote points in  $\Omega$  and  $x = (\underline{x}, x_3) = (x_1, x_2, x_3)$  to denote points in  $Q^- = \Omega \times (-h/2, 0)$  or  $Q^+ = \Omega \times (0, h/2)$ . For  $x \in Q^+ \cup Q^-$  let  $U(x) = (U_1, U_2, U_3)(x)$  denote the displacement vector of the point which, when the plate in equilibrium has coordinates  $x = (x_1, x_2, x_3)$ . (In the dynamic case these variables also depend upon time, but we will suppress all time dependence in this section.) In addition let us define  $u^+$  and  $u^-$  by

$$u^\pm(\underline{x}) = \lim_{x_3 \rightarrow 0^\pm} U(\underline{x}, x_3) \quad \forall \underline{x} \in \Omega.$$

**2.1. Stress-strain relations.** Let  $\sigma_{ij}, \epsilon_{ij}$  ( $i, j = 1, 2, 3$ ) denote the stress and strain tensors, respectively. For a small displacement theory we assume

$$(2.1) \quad \epsilon_{ij} = \frac{1}{2} \left( \frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} \right).$$

Let  $E = E(\underline{x}) > E_0 > 0$  denote Young's modulus and  $\mu(\underline{x})$  denote Poisson's ratio.  $0 < \mu_0 < \mu < \mu_1 < 1/2$ . Since the plates are assumed to be isotropic we have the following stress-strain relations:

$$(2.2) \quad \sigma_{ij} = \frac{E}{1+\mu} \left( \epsilon_{ij} + \frac{\mu}{1-2\mu} \epsilon_{kk} \delta_{ij} \right),$$

where we have used the summation convention for repeated indices.

As in [LL], we assume that  $\sigma_{33}$  is negligible, hence to highest order one can solve for  $\epsilon_{33}$  in terms of the other principle strains:

$$(2.3) \quad \epsilon_{33} = \frac{-\mu}{1-\mu} (\epsilon_{11} + \epsilon_{22}).$$

From (2.2) and (2.3) we have

$$(2.4) \quad \begin{cases} \sigma_{11} &= \frac{E}{1-\mu^2} (\epsilon_{11} + \mu \epsilon_{22}) \\ \sigma_{22} &= \frac{E}{1-\mu^2} (\mu \epsilon_{11} + \epsilon_{22}) \\ \sigma_{33} &= 0 \\ \sigma_{ij} &= \frac{E}{1+\mu} \epsilon_{ij}, i \neq j. \end{cases}$$

However, in anticipation of the fact that we will model each plate according to the assumptions of Timoshenko beam theory, a *shear correction* coefficient  $k$  is incorporated into (2.4) to account for the fact that (physically) the transverse shear strains do not remain constant throughout the thickness of each plate. (In the Timoshenko theory the shear strains are assumed to be constant throughout the thickness.) Thus the shear stresses in (2.4) are modified:

$$(2.5) \quad \sigma_{13} = \frac{kE}{1+\mu} \epsilon_{13}, \quad \sigma_{23} = \frac{kE}{1+\mu} \epsilon_{23}.$$

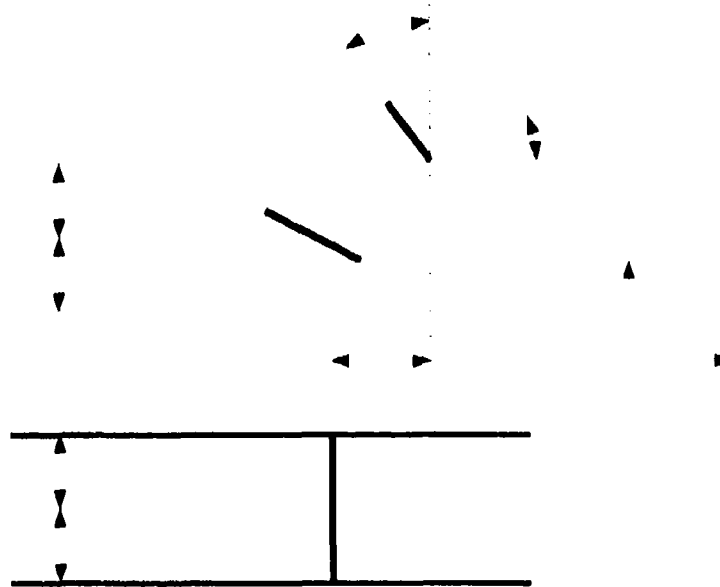


FIG. 1. A schematic of the  $x_1$ - $x_3$  cross section of the undeformed filament  $A$ - $B$ - $C$ - $D$  (bottom) located at  $(\underline{x}, \cdot)$  and the corresponding deformed filaments  $A'$ - $B'$ ,  $C'$ - $D'$  (top) whose positions are described by  $\psi_i^\pm(\underline{x})$ ,  $\theta_i^\pm(\underline{x})$ ,  $u_i^\pm(\underline{x})$ .

**2.2. Plate assumptions.** The MTR plate model [LL, p.14] is the simplest plate model which incorporates the effects of shear as an independent quantity. In this model straight filaments originally orthogonal to a "middle surface" remain straight after deformation, however are free to rotate relative to the deformed "middle surface". Thus the MTR system is described by three equations: one for the dynamics of the transverse deflection and two for the shear angles. If in addition, one allows for in-plane translations, one obtains an additional two equations for the dynamics of these motions.

Both the upper and lower plates of our composite plate will be modelled by MTR plates which include the in-plane translations. Thus within each plate, the straight filaments are allowed both rotate and translate so that each filament in each plate has (for the moment) five degrees of freedom: three translational and two rotational. Thus the displacements of the filament  $\mathcal{F} = \mathcal{F}^+ \cup \mathcal{F}^-$  which occupies  $\{(x_1, x_2, x_3) : -h/2 < x_3 < h/2\}$  at equilibrium is completely determined by specifying the translational displacements  $u^\pm(\underline{x})$  and a rotation angle for  $\mathcal{F}^+$  and  $\mathcal{F}^-$ . ( $\mathcal{F}^+$  belongs to the "upper plate" and  $\mathcal{F}^-$  is in the "lower plate".)

Rather than attempt to define shear angles relative to a "middle surface" of each plate, it will be more convenient to define the angles relative to the surfaces  $\mathcal{S}^\pm = \{(\underline{x}, 0) + u^\pm(\underline{x}) : \underline{x} \in \Omega\}$ . Let  $\varphi^\pm(\underline{x}) = (\varphi_1^\pm, \varphi_2^\pm)(\underline{x})$  denote shear angles; that

is,  $\varphi_i^\pm(\underline{x})$  is the angle, measured with positive orientation that the filament  $\mathcal{F}^\pm$  has rotated relative to the normal to  $\mathcal{S}^\pm$  at  $(\underline{x}, 0) + u^\pm(\underline{x})$ . Let  $\theta^\pm(\underline{x}) = (\theta_1^\pm, \theta_2^\pm)(\underline{x})$  denote the rotation angles from the equilibrium position, with positive orientation, of  $\mathcal{S}^\pm$  at  $(\underline{x}, 0) + u^\pm(\underline{x})$ . Let  $\psi^\pm(\underline{x}) = (\psi_1^\pm, \psi_2^\pm)(\underline{x})$  denote the total rotation angle the filament  $\mathcal{F}^\pm$  has rotated with respect to the  $x_3$  axis, again, with positive orientation. (See Fig. 1.) The angles  $\theta^\pm$ ,  $\psi^\pm$  and  $\varphi^\pm$  are thus related by

$$(2.6) \quad \psi^\pm(\underline{x}) = \theta^\pm(\underline{x}) + \varphi^\pm(\underline{x}) \quad \forall \underline{x} \in \Omega.$$

Under the assumption that the rotation angles of the middle surface are small we may use the approximation

$$(2.7) \quad \theta^\pm(\underline{x}) \simeq \nabla u_3^\pm(\underline{x}) \quad \forall \underline{x} \in \Omega.$$

Then the displacements can be written in terms of the translations  $u^\pm$  and total rotation angles  $\psi^\pm$  as

$$(2.8) \quad \begin{cases} U_1(x_1, x_2, x_3) = \begin{cases} u_1^+(x_1, x_2) - x_3 \psi_1^+(x_1, x_2) & x_3 > 0 \\ u_1^-(x_1, x_2) - x_3 \psi_1^-(x_1, x_2) & x_3 < 0 \end{cases} \\ U_2(x_1, x_2, x_3) = \begin{cases} u_2^+(x_1, x_2) - x_3 \psi_2^+(x_1, x_2) & x_3 > 0 \\ u_2^-(x_1, x_2) - x_3 \psi_2^-(x_1, x_2) & x_3 < 0 \end{cases} \\ U_3(x_1, x_2, x_3) = \begin{cases} u_3^+(x_1, x_2) & x_3 > 0 \\ u_3^-(x_1, x_2) & x_3 < 0 \end{cases} \end{cases}$$

Substituting (2.8) into (2.1), and using (2.6), (2.7) gives

$$(2.9) \quad \begin{cases} \epsilon_{11} = \frac{\partial u_1^\pm}{\partial x_1} - x_3 \frac{\partial \psi_1^\pm}{\partial x_1} & \epsilon_{22} = \frac{\partial u_2^\pm}{\partial x_2} - x_3 \frac{\partial \psi_2^\pm}{\partial x_2} \\ \epsilon_{12} = \frac{1}{2} \left[ \frac{\partial u_1^\pm}{\partial x_2} + \frac{\partial u_2^\pm}{\partial x_1} - x_3 \left( \frac{\partial \psi_1^\pm}{\partial x_2} + \frac{\partial \psi_2^\pm}{\partial x_1} \right) \right] \\ \epsilon_{13} = \frac{1}{2}(-\varphi_1^\pm) & \epsilon_{23} = \frac{1}{2}(-\varphi_2^\pm), \end{cases}$$

where the + superscript (respectively, - subscript) applies to  $x_3 > 0$  (respectively,  $x_3 < 0$ ). Instead of using (2.3) however, we set (in accordance with [LL, p. 9])

$$(2.10) \quad \epsilon_{33} = 0.$$

One can check that (2.8)-(2.9) is the exact analogue of the strain-displacement equations which one obtains for the MTR system as in [LL, p.14]. However up to this point no assumptions have been made regarding the amount of slip between the two plates.

**2.3. Interface assumptions.** For  $\underline{x} \in \Omega$  let us define a dimensionless slip variable  $s(\underline{x}) = (s_1, s_2)(\underline{x})$  by

$$(2.11) \quad s_i(\underline{x}) = \frac{u_i^+(\underline{x}) - u_i^-(\underline{x})}{2h} \quad i = 1, 2.$$

For the moment consider the glue layer to have a uniform thickness  $d$  and that shearing motions obey Hooke's law. In this case the slip  $s(\underline{x})$  is proportional to the amount of shear at  $\underline{x}$  and thus integrating over  $\Omega$  the product of the stress and the strain gives the following expression for the strain energy within the glue:

$$(2.12) \quad \mathcal{P}_g = \frac{h^2}{2} \int_{\Omega} \gamma(s_1^2(\underline{x}) + s_2^2(\underline{x})) d\underline{x},$$

where  $\gamma = \gamma(\underline{x}) \geq 0$  is proportional to  $d$  and the shear modulus of the glue.

Henceforth we shall regard (2.12) as the only (non-negligible) contribution by the glue to energy for the composite plate, where  $\gamma$  is an arbitrary non-negative essentially bounded function defined on  $\bar{\Omega}$ . Thus implicitly we assume that the mass and thickness of the "glue" which bonds the surfaces is negligible (compared to those of the plates) and can be ignored. While the argument used to motivate the assumption (2.12) is heuristic, (2.12) provides the simplest possible means by which the slip can be penalized.

Let

$$w(\underline{x}) = \frac{u_3^+(\underline{x}) + u_3^-(\underline{x})}{2}, \quad \theta(\underline{x}) = \frac{\theta^+(\underline{x}) + \theta^-(\underline{x})}{2}, \quad \underline{x} \in \Omega.$$

To highest order, Taylor's formula gives (we will use the usual "dot product" notation):

$$(2.13) \quad u_3^\pm(\underline{x}) - w(\underline{x}) = \pm h \nabla w(\underline{x}) \cdot s(\underline{x}) \quad \underline{x} \in \Omega.$$

To obtain a linear theory, we assume the time and spatial derivatives of the right hand side of (2.13) are negligible in comparison to those of  $w$ . Thus let us assume

$$(2.14) \quad w(\underline{x}) = u_3^+(\underline{x}) = u_3^-(\underline{x}).$$

In (2.14) we are actually making the same assumption that is made in the MTR theory, namely that the transverse displacements on the deformed filaments  $\mathcal{F}$  (described earlier in this section) are independent of  $x_3$ .

**2.4. Strain and kinetic energy.** The strain energy for the composite plate consists of the strain energy for the glue (2.12) and the strain energy for the plates

$$(2.15) \quad \mathcal{P}^+ = \frac{1}{2} \int_{Q^+} \epsilon_{ij} \sigma_{ij} d\underline{x} dx_3.$$



From (2.4) and (2.5) we can write  $\mathcal{P}_{\pm}$  in terms of the strains:

$$(2.16) \quad \mathcal{P}^- + \mathcal{P}^+ = \frac{1}{2} \int_{Q^+ \cup Q^-} \frac{E}{(1-\mu^2)} (\epsilon_{11}^2 + 2\mu\epsilon_{11}\epsilon_{22} + \epsilon_{22}^2 + 2(1-\mu)(k\epsilon_{12}^2 + k\epsilon_{13}^2 + \epsilon_{23}^2)) d\mathbf{x} dx_3.$$

For  $\mathbf{x} \in \Omega$  define  $u = (u_1, u_2)$ ,  $\psi = (\psi_1, \psi_2)$ ,  $\eta = (\eta_1, \eta_2)$  by

$$(2.17) \quad \begin{cases} u_i(\mathbf{x}) &= \frac{1}{2}(u_i^+(\mathbf{x}) + u_i^-(\mathbf{x})) \quad i = 1, 2 \\ \psi(\mathbf{x}) &= \frac{1}{2}(\psi^+(\mathbf{x}) + \psi^-(\mathbf{x})) \\ \eta(\mathbf{x}) &= \frac{1}{2}(\psi^+(\mathbf{x}) - \psi^-(\mathbf{x})). \end{cases}$$

Then we have

$$(2.18) \quad \begin{cases} u_i^+(\mathbf{x}) &= u_i(\mathbf{x}) + h s_i(\mathbf{x}) \quad i = 1, 2 \\ u_i^-(\mathbf{x}) &= u_i(\mathbf{x}) - h s_i(\mathbf{x}) \quad i = 1, 2 \\ \psi^+(\mathbf{x}) &= \psi(\mathbf{x}) + \eta(\mathbf{x}) \\ \psi^-(\mathbf{x}) &= \psi(\mathbf{x}) - \eta(\mathbf{x}). \end{cases}$$

Substitution of (2.18), (2.9), (2.10), (2.14), into (2.16) gives an expression for  $\mathcal{P}^+ + \mathcal{P}^-$  in terms of  $s, u, \psi$  and  $\eta$ . For example,

$$\begin{aligned} \int_{-h/2}^{h/2} \epsilon_{11}^2 dx_3 &= 2 \int_0^{h/2} \left( \frac{\partial u_1}{\partial x_1} \right)^2 + h^2 \left( \frac{\partial s_1}{\partial x_1} \right)^2 + x_3^2 \left( \left( \frac{\partial \psi_1}{\partial x_1} \right)^2 + \left( \frac{\partial \eta_1}{\partial x_1} \right)^2 \right) \\ &\quad - 2x_3 \left( \frac{\partial u_1}{\partial x_1} \frac{\partial \eta_1}{\partial x_1} + \frac{\partial \psi_1}{\partial x_1} h \frac{\partial s_1}{\partial x_1} \right) dx_3 \\ &= \frac{h^3}{12} \left[ \left( \frac{\partial}{\partial x_1} (\psi_1 - 3s_1) \right)^2 + 3 \left( \frac{\partial s_1}{\partial x_1} \right)^2 \right] \\ &\quad + \frac{h}{12} \left[ \left( \frac{\partial}{\partial x_1} (h\eta_1 - 3u_1) \right)^2 + 3 \left( \frac{\partial u_1}{\partial x_1} \right)^2 \right]. \end{aligned}$$

Let

$$(2.19) \quad \xi(\mathbf{x}) = \psi(\mathbf{x}) - 3s(\mathbf{x}), \quad \mathbf{x} \in \Omega$$

$$(2.20) \quad \nu(\mathbf{x}) = h\eta(\mathbf{x}) - 3u(\mathbf{x}), \quad \mathbf{x} \in \Omega.$$

Also let us define an average shear angle  $\varphi = (\varphi_1, \varphi_2)$  by

$$(2.21) \quad \varphi(\mathbf{x}) = \psi(\mathbf{x}) - \nabla w(\mathbf{x}) = \xi(\mathbf{x}) + 3s(\mathbf{x}) - \nabla w(\mathbf{x}).$$

When each of the terms in (2.16) are computed one finds that the potential energy of the plates splits into a *bending potential energy* and a *stretching potential energy*:

$$(2.22) \quad \mathcal{P}^- + \mathcal{P}^+ = \mathcal{P}_b + \mathcal{P}_s,$$

where

$$(2.23) \quad \mathcal{P}_b = \frac{1}{2} \int_{\Omega} D \left[ \left( \frac{\partial \xi_i}{\partial x_i} \right)^2 + 3 \left( \frac{\partial s_i}{\partial x_i} \right)^2 + 2\mu \left( \frac{\partial \xi_2}{\partial x_2} \frac{\partial \xi_1}{\partial x_1} + 3 \frac{\partial s_1}{\partial x_1} \frac{\partial s_2}{\partial x_2} \right) \right. \\ \left. + \left( \frac{1-\mu}{2} \right) \left[ \left( \frac{\partial \xi_1}{\partial x_2} + \frac{\partial \xi_2}{\partial x_1} \right)^2 + 3 \left( \frac{\partial s_1}{\partial x_2} + \frac{\partial s_2}{\partial x_1} \right)^2 \right] \right] + K \varphi \cdot \varphi d\underline{x}$$

and

$$\mathcal{P}_s = \frac{1}{2} \int_{\Omega} \frac{Eh}{12(1-\mu^2)} \left[ \left( \frac{\partial \nu_i}{\partial x_i} \right)^2 + 3 \left( \frac{\partial u_i}{\partial x_i} \right)^2 + 2\mu \left( \frac{\partial \nu_2}{\partial x_2} \frac{\partial \nu_1}{\partial x_1} + 3 \frac{\partial u_1}{\partial x_1} \frac{\partial u_2}{\partial x_2} \right) \right. \\ \left. + \left( \frac{1-\mu}{2} \right) \left[ \left( \frac{\partial \nu_1}{\partial x_2} + \frac{\partial \nu_2}{\partial x_1} \right)^2 + 3 \left( \frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right)^2 \right] \right] d\underline{x}$$

where  $D = Eh^3/12(1-\mu^2)$  is the *modulus of flexural rigidity* (see [LL, p. 10]) for a plate of thickness  $h$ , with the same material properties as the two plates (which each have thickness  $h/2$ ), and  $K = kEh/2(1+\mu)$  is the *modulus of elasticity in shear* [LL, p. 14], also for the same plate, but with thickness  $h$ .

The kinetic energy  $\mathcal{K} = \mathcal{K}^+ + \mathcal{K}^-$  is defined by

$$\mathcal{K}_{\pm} = \frac{1}{2} \int_{Q^{\pm}} \rho (\dot{U}_1^2 + \dot{U}_2^2 + \dot{U}_3^2) d\underline{x} dx_3,$$

where  $\dot{\phantom{x}} = d/dt$  and  $\rho = \rho(\underline{x}) > \rho_0 > 0$  denotes the mass density per unit volume.

As with the potential energy, the kinetic energy decouples into a bending part and a stretching part:

$$(2.24) \quad \mathcal{K} = \mathcal{K}_b + \mathcal{K}_s.$$

One finds that

$$(2.25) \quad \mathcal{K}_b = \frac{1}{2} \int_{\Omega} \rho h (\dot{w})^2 + I_{\rho} \left( \dot{\xi} \cdot \dot{\xi} + 3 \dot{s} \cdot \dot{s} \right) d\underline{x},$$

$$(2.26) \quad \mathcal{K}_s = \frac{1}{2} \int_{\Omega} \frac{\rho h}{12} (\dot{\nu} \cdot \dot{\nu} + 3 \dot{u} \cdot \dot{u}) d\underline{x},$$

where  $I_{\rho} = \rho h^3/12$  is the density function for the mass-moment of inertia about the interface.

**2.5. Work.** Assume that  $\Omega$  has smooth boundary  $\Gamma$ . To set ideas we will assume that the plate is clamped on a portion of its edge  $\Gamma_0 \subset \Gamma$ . Furthermore denote  $\Gamma_1 = \Gamma - \Gamma_0$ .

Now assume the composite plate is subject to a volume distribution of forces  $(\tilde{f}_1, \tilde{f}_2, \tilde{f}_3)$  and a distribution of forces  $(\tilde{g}_1, \tilde{g}_2, \tilde{g}_3)$  along  $\Gamma_1$ . The work done on the plate by these forces is

$$\mathcal{W} = \int_{-h/2}^{h/2} \int_{\Omega} \tilde{f}_i U_i d\underline{x} dx_3 + \int_{-h/2}^{h/2} \int_{\Gamma_1} \tilde{g}_i U_i d\Gamma dx_3.$$

It will not be necessary to consider the pointwise forces, but only the resultants

$$(2.27) \quad \begin{cases} f_i(\underline{x}) = \int_{-h/2}^{h/2} \tilde{f}_i d\underline{x} dx_3 & g_i(\underline{x}) = \int_{-h/2}^{h/2} \tilde{g}_i d\underline{x} dx_3, \quad i = 1, 2, 3 \\ M_i(\underline{x}) = \int_{-h/2}^{h/2} \tilde{f}_i x_3 d\underline{x} dx_3 & m_i(\underline{x}) = \int_{-h/2}^{h/2} \tilde{g}_i x_3 d\underline{x} dx_3 \quad i = 1, 2 \\ P_i(\underline{x}) = \int_{-h/2}^{h/2} \tilde{f}_i \operatorname{sgn}(x_3) d\underline{x} dx_3 & p_i(\underline{x}) = \int_{-h/2}^{h/2} \tilde{g}_i \operatorname{sgn}(x_3) d\underline{x} dx_3 \quad i = 1, 2 \\ Q_i(\underline{x}) = \int_{-h/2}^{h/2} \tilde{f}_i |x_3| d\underline{x} dx_3 & q_i(\underline{x}) = \int_{-h/2}^{h/2} \tilde{g}_i |x_3| d\underline{x} dx_3 \quad i = 1, 2. \end{cases}$$

As with the energy expressions,  $\mathcal{W}$  decouples into two parts

$$(2.28) \quad \mathcal{W} = \mathcal{W}_s + \mathcal{W}_b;$$

$$(2.29) \quad \begin{aligned} \mathcal{W}_b = & \int_{\Omega} h s_1 P_1 + h s_2 P_2 - \psi_1 M_1 - \psi_2 M_2 + w f_3 d\underline{x} \\ & + \int_{\Gamma_1} h s_1 p_1 + h s_2 p_2 - \psi_1 m_1 - \psi_2 m_2 + w g_3 d\Gamma. \end{aligned}$$

$$(2.30) \quad \begin{aligned} \mathcal{W}_s = & \int_{\Omega} u_1 f_1 + u_2 f_2 - \eta_1 Q_1 - \eta_2 Q_2 d\underline{x} \\ & + \int_{\Gamma_1} u_1 g_1 + u_2 g_2 - \eta_1 q_1 - \eta_2 q_2 d\Gamma. \end{aligned}$$

### 3. Equations of motion

The Lagrangian  $\mathcal{L}$  on  $(0, T)$  is defined by

$$\mathcal{L} = \int_0^T \mathcal{K}(t) + \mathcal{W}(t) - \mathcal{P}(t) dt.$$

Because of the uncoupling in (2.22), (2.24) and (2.28) of the variables  $(w, s, \psi)$  from  $(u, \eta)$  in the expressions for  $\mathcal{K}$ ,  $\mathcal{W}$  and  $P$ , the Lagrangian also decouples as  $\mathcal{L} = \mathcal{L}_b + \mathcal{L}_s$ , where

$$(3.1) \quad \mathcal{L}_b = \int_0^T \mathcal{K}_b(t) + \mathcal{W}_b(t) - \mathcal{P}_b(t) - \mathcal{P}_g(t) dt$$

$$(3.2) \quad \mathcal{L}_s = \int_0^T \mathcal{K}_s(t) + \mathcal{W}_s(t) - \mathcal{P}_s(t) dt.$$

**3.1. Weak form of equations.** According to the principle of virtual work, the solution trajectory is the trajectory which renders stationary the Lagrangian under all kinematically admissible displacements. Due to the decoupling of (3.1), (3.2) this will hold if and only if the variations of  $\mathcal{L}_b$  and  $\mathcal{L}_s$  vanish separately. Thus the equations of motion one obtains decouples into two *completely independent parts*. By setting the variation of  $\mathcal{L}_s = 0$  one obtains a system of equations in the variables  $u$  and  $\eta$  which describe the in-plane or *stretching* motions. We mention what these are in Remark 3.3.

On the other hand, the equations one obtains from  $\mathcal{L}_b$  describe *bending* motions. Since our primary motivation is to study the effect that slip has upon bending we are mainly interested in the equations which one obtains from  $\mathcal{L}_b$ .

It thus follows that for the purpose of calculating  $(\xi, s, w)$  we may assume *with no loss of generality* that

(H1) Solutions satisfy

$$(i) \quad \psi^+ = \psi^- = \psi$$

$$(ii) \quad u^+ + u^- = 0.$$

(H2) All terms in  $\mathcal{W}_s$  vanish, i.e.

$$(i) \quad f_1 = f_2 = g_1 = g_2 = 0$$

$$(ii) \quad Q_1 = Q_2 = q_1 = q_2 = 0.$$

Let  $(\hat{\xi}, \hat{s}, \hat{w})$  denote a test function on  $\Omega \times (0, T)$  for which

$$(\hat{\xi}, \hat{s}, \hat{w}) = \left( \frac{\partial \hat{\xi}}{\partial n}, \frac{\partial \hat{s}}{\partial n}, \frac{\partial \hat{w}}{\partial n} \right) = 0 \quad \text{on } \Gamma_0 \times (0, T)$$

$$(\hat{\xi}, \hat{s}, \hat{w})|_{t=0} = \frac{\partial}{\partial t} (\hat{\xi}, \hat{s}, \hat{w})|_{t=0} = (\hat{\xi}, \hat{s}, \hat{w})|_{t=T} = \frac{\partial}{\partial t} (\hat{\xi}, \hat{s}, \hat{w})|_{t=T} = 0 \quad \text{in } \Omega$$

where  $n$  is the outward unit normal to  $\Gamma$ . We set

$$0 = \langle \mathcal{L}'_b(\xi, s, w), (\hat{\xi}, \hat{s}, \hat{w}) \rangle = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{L}_b((\xi, s, w) + \epsilon(\hat{\xi}, \hat{s}, \hat{w})) - \mathcal{L}_b(\xi, s, w)}{\epsilon}$$

to obtain the equations of motion in weak form:

$$(3.3) \quad \int_0^T c(\dot{\xi}, \dot{s}, \dot{w}; \dot{\xi}, \dot{s}, \dot{w}) - a(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) + (f_3, \dot{w})_\Omega \\ - (M, \dot{\psi})_\Omega + h(P, \dot{s})_\Omega dt + \int_0^T \int_{\Gamma_1} hp \cdot \dot{s} - m \cdot \dot{\psi} + \dot{w} g_3 d\Gamma dt = 0$$

where  $(f, g)_\Omega = \int_\Omega f g d\Omega$ ,  $\dot{\psi} = \dot{\xi} + 3\dot{s}$  and

$$(3.4) \quad c(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) = h(\rho w, \dot{w})_\Omega + (I_\rho \xi, \dot{\xi})_\Omega + (3I_\rho s, \dot{s})_\Omega$$

$$(3.5) \quad a(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) = a_0(\xi; \dot{\xi}) + 3a_0(s; \dot{s}) + a_1(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) + a_2(s; \dot{s})$$

$$(3.6) \quad a_0(\xi, \dot{\xi}) = \left( D \frac{\partial \xi_1}{\partial x_1}, \frac{\partial \dot{\xi}_1}{\partial x_1} \right)_\Omega + \left( D \frac{\partial \xi_2}{\partial x_2}, \frac{\partial \dot{\xi}_2}{\partial x_2} \right)_\Omega \\ + \left( \mu D \frac{\partial \xi_2}{\partial x_2}, \frac{\partial \dot{\xi}_1}{\partial x_1} \right)_\Omega + \left( \mu D \frac{\partial \xi_1}{\partial x_1}, \frac{\partial \dot{\xi}_2}{\partial x_2} \right)_\Omega \\ + \left( \left( \frac{1-\mu}{2} \right) D \left( \frac{\partial \xi_1}{\partial x_2} + \frac{\partial \xi_2}{\partial x_1} \right), \left( \frac{\partial \dot{\xi}_1}{\partial x_2} + \frac{\partial \dot{\xi}_2}{\partial x_1} \right) \right)_\Omega$$

$$(3.7) \quad a_1(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) = (K(\xi + 3s - \nabla w), \dot{\xi} + 3\dot{s} - \nabla \dot{w})_\Omega$$

$$(3.8) \quad a_2(s, \dot{s}) = 4h^2(\gamma s, \dot{s})_\Omega.$$

The force terms in (3.3) can be rewritten

$$(3.9) \quad (hP, \dot{s})_\Omega - (M, \dot{\psi})_\Omega = -(M, \dot{\xi})_\Omega + (\mathcal{K}, \dot{s})_\Omega; \quad \mathcal{K} = hP - 3M$$

and

$$(3.10) \quad \int_{\Gamma_1} hp \cdot \dot{s} - m \cdot \dot{\psi} d\mu = \int_{\Gamma_1} -m \cdot \dot{\xi} + \kappa \cdot \dot{s}; \quad \kappa = hp - 3m.$$

**3.2. Associated boundary value problem.** The boundary value problem associated with (3.3) can be obtained through the usual integration by parts procedure. Before we state the result let us first introduce some simplifying notation.

If  $\varepsilon = (\varepsilon)_{ij}$  denotes a symmetric  $2 \times 2$  matrix then  $M[\varepsilon]$  defined by

$$(3.11) \quad M[\varepsilon] = D \begin{pmatrix} \varepsilon_{11} + \mu \varepsilon_{22} & (1-\mu)\varepsilon_{12} \\ (1-\mu)\varepsilon_{12} & \mu \varepsilon_{11} + \varepsilon_{22} \end{pmatrix},$$

defines another symmetric matrix. For differentiable  $\psi = (\psi_1, \psi_2)$  define  $\epsilon(\psi)$  by

$$(3.12) \quad \epsilon(\psi) = \frac{1}{2}(\nabla\psi + (\nabla\psi)^T) = \frac{1}{2} \left( \frac{\partial\psi_i}{\partial x_j} + \frac{\partial\psi_j}{\partial x_i} \right)_{ij}.$$

Then (3.11)-(3.12) defines a symmetric  $2 \times 2$  matrix  $\mathcal{M}[z]$  by

$$(3.13) \quad \mathcal{M}[z] = M[\epsilon(z)].$$

We define the divergence of a symmetric matrix to be the divergence of each row:

$$\operatorname{div} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} = (\operatorname{div}(a_{11}, a_{12}), \operatorname{div}(a_{12}, a_{22})).$$

Then  $L\psi = (L_1\psi, L_2\psi) = \operatorname{div} \mathcal{M}[\psi]$  defines a second order operator which is given explicitly by

$$(3.14) \quad \begin{aligned} L_i\psi &= \frac{\partial}{\partial x_i} \left( D \frac{\partial\psi_i}{\partial x_i} \right) + \frac{\partial}{\partial x_j} \left( \frac{(1-\mu)}{2} D \frac{\partial\psi_i}{\partial x_j} \right) + \frac{\partial}{\partial x_j} \left( \frac{(1-\mu)}{2} D \frac{\partial\psi_j}{\partial x_i} \right) \\ &\quad + \frac{\partial}{\partial x_i} \left( \mu D \frac{\partial\psi_j}{\partial x_j} \right), \quad (i, j) = (1, 2), (2, 1). \end{aligned}$$

Let us also define the boundary operator  $\mathcal{B}\psi = (\mathcal{B}_1(\psi_1, \psi_2), (\mathcal{B}_2(\psi_1, \psi_2)))$  by

$$(3.15) \quad \mathcal{B}\psi = \mathcal{M}[\psi]n,$$

where  $n = (n_1, n_2)$  denotes the outward unit normal to  $\Gamma$ . Explicitly one has

$$\begin{aligned} \mathcal{B}_1(\psi_1, \psi_2) &= D \left[ \left( \frac{\partial\psi_1}{\partial x_1} n_1 + \mu \frac{\partial\psi_2}{\partial x_2} n_1 \right) + \left( \frac{1-\mu}{2} \right) \left( \frac{\partial\psi_1}{\partial x_2} + \frac{\partial\psi_2}{\partial x_1} \right) n_2 \right] \\ \mathcal{B}_2(\psi_1, \psi_2) &= D \left[ \left( \frac{\partial\psi_2}{\partial x_2} n_2 + \mu \frac{\partial\psi_1}{\partial x_1} n_2 \right) + \left( \frac{1-\mu}{2} \right) \left( \frac{\partial\psi_2}{\partial x_1} + \frac{\partial\psi_1}{\partial x_2} \right) n_1 \right]. \end{aligned}$$

An integration by parts in  $t$  of (3.3) followed by an application of Green's theorem leads to the associated boundary value problem:

$$(3.16) \quad \begin{cases} \text{(i)} & \rho h \ddot{w} + \operatorname{div}(K\varphi) = f_3 & \text{in } \Omega \times \mathbb{R}^+ \\ \text{(ii)} & I_\rho \ddot{\xi} + K\varphi - L\xi = -M & \text{in } \Omega \times \mathbb{R}^+ \\ \text{(iii)} & 3I_\rho \ddot{s} + 3K\varphi + 4\gamma h^2 s - 3Ls = \mathcal{K} & \text{in } \Omega \times \mathbb{R}^+ \\ \text{(iv)} & \varphi = \xi + 3s - \nabla w & \text{in } \Omega \times \mathbb{R}^+ \end{cases}$$

with the boundary conditions

$$(3.17) \quad \begin{cases} \text{(i)} & (w, \xi, s) = 0 & \text{on } \Gamma_0 \times \mathbb{R}^+ \\ \text{(ii)} & K\varphi \cdot n = g_3 & \text{on } \Gamma_1 \times \mathbb{R}^+ \\ \text{(iii)} & \mathcal{B}\xi = -m & \text{on } \Gamma_1 \times \mathbb{R}^+ \\ \text{(iv)} & 3\mathcal{B}s = \kappa & \text{on } \Gamma_1 \times \mathbb{R}^+. \end{cases}$$

Initial conditions can be given as

$$(3.18) \quad \begin{cases} \xi_i^0(\underline{x}) = \xi_i(\underline{x}, 0); \quad \xi_i^1(\underline{x}) = \dot{\xi}_i(\underline{x}, 0) & i = 1, 2 \\ s_i^0(\underline{x}) = s_i(\underline{x}, 0); \quad s_i^1(\underline{x}) = \dot{s}_i(\underline{x}, 0) & i = 1, 2 \\ w^0(\underline{x}) = w(\underline{x}, 0); \quad w^1(\underline{x}) = \dot{w}(\underline{x}, 0) & \underline{x} \in \Omega. \end{cases}$$

Of course, depending upon the regularity of the initial data the solution of (3.16)-(3.17) will have different regularity properties.

Many other types of boundary conditions are possible, and easily determined by inspection of (3.17). In the boundary conditions we consider, (i) is precisely the usual "clamped" b.c.'s together with the extra condition that the slip on  $\Gamma_0$  vanishes, (ii) specifies the transverse shear on  $\Gamma_1$  while (iii) and (iv) are moment conditions. It's worth noting that the moments  $\mathcal{K}$  and  $\kappa$  are the unique ones (up to multiplicity) for which  $M \cdot \mathcal{K} = 0$  at each  $\underline{x} \in \Omega$  and  $m \cdot \kappa = 0$  at each  $t \in \Gamma_1$ .

*Remark 3.1.* The natural substitution  $\xi = \psi - 3s$  used in (2.19) suggests that  $s$  has an interpretation as an angle. This is justified by noting that to highest order  $\xi_i$  is the rotation angle (projected onto the  $x_i$ - $x_3$  plane with positive orientation as measured from the  $x_3$  axis) of the ray from  $(0, 0, 0)$  to  $(U_1(\cdot, \cdot, h/3), U_2(\cdot, \cdot, h/3), U_3(\cdot, \cdot, h/3))$ . Thus  $\xi$  can be thought of as an *effective angle of rotation*.

*Remark 3.2.* The fact that the Lagrangian decouples into a bending part and a stretching part relied upon finding the change of variables (2.11), (2.17). If however, we did not assume that the top and bottom plates were identical, then except in very special cases, there does not exist a change of variables for which the Lagrangian decouples. Said another way, without the symmetry in the upper and lower plates, the in-plane motions *do not decouple* from the transverse motions.

*Remark 3.3.* In the same way that (3.16)-(3.17) were obtained, one may obtain the equations for the in-plane motions. In the following, the  $u$  variables represent the coordinates of the displacement of the center of mass while the  $\nu$  variables are related to the difference in the shear angles of the top and bottom plate (see (2.17), (2.20)).

$$\begin{cases} \text{(i)} & I_\rho \ddot{\nu} - L\nu = -Qh & \text{in } \Omega \times \mathbb{R}^+ \\ \text{(ii)} & I_\rho \ddot{u} - Lu = h^2(f_1, f_2) - 3Qh & \text{in } \Omega \times \mathbb{R}^+ \end{cases}$$

with the boundary conditions

$$\begin{cases} \text{(i)} & \nu = u = 0 & \text{on } \Gamma_0 \times \mathbb{R}^+ \\ \text{(ii)} & \mathcal{B}\nu = -qh & \text{on } \Gamma_1 \times \mathbb{R}^+ \\ \text{(iii)} & \mathcal{B}u = h^2(g_1, g_2) - 3hq & \text{on } \Gamma_1 \times \mathbb{R}^+. \end{cases}$$

#### 4. Existence, uniqueness, regularity

Denote

$$H_{\Gamma_0}^1 = \{\varphi : \varphi \in H^1(\Omega), \varphi = 0 \text{ on } \Gamma_0\}.$$

Recall  $\Gamma_0$  could be part or all of  $\Gamma$ . In the latter case  $H_{\Gamma_0}^1 = H_0^1(\Omega)$ . Also define

$$\mathcal{V} = (H_{\Gamma_0}^1)^5 \quad \mathcal{H} = (L^2(\Omega))^5.$$

**4.1. Variational formulation.** Let us consider our problem in the absence of external forces. In this case the variational formulation of (3.16)-(3.18) is: Find functions  $\xi, s, w$  such that

$$(4.1) \quad (\xi, s, w) \in C^1([0, T]; \mathcal{V}) \cap C^1([0, T]; \mathcal{H}),$$

$$(4.2) \quad \frac{d}{dt} c(\dot{\xi}, \dot{s}, \dot{w}; \hat{\xi}, \hat{s}, \hat{w}) + a(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) = 0 \quad \forall (\dot{\xi}, \dot{s}, \dot{w}) \in \mathcal{V}$$

in the sense of distributions on  $(0, T)$  and

$$(4.3) \quad \begin{cases} (\xi(0), s(0), w(0)) = (\xi^0, s^0, w^0) & \text{given in } \mathcal{V} \\ (\dot{\xi}(0), \dot{s}(0), \dot{w}(0)) = (\xi^1, s^1, w^1) & \text{given in } \mathcal{H}. \end{cases}$$

The forms  $c(\cdot : \cdot)$ ,  $a(\cdot : \cdot)$ ,  $a_0(\cdot : \cdot)$ ,  $a_1(\cdot : \cdot)$  and  $a_2(\cdot : \cdot)$  are each symmetric bilinear functions of their arguments. Let us denote by  $c(\cdot)$ ,  $a(\cdot)$  and so forth the corresponding (non-negative) quadratic functions, e.g.,  $a(\xi, s, w) = a(\xi, s, w; \xi, s, w)$ .

**Theorem 4.1.** *The variational problem (4.1)-(4.3) admits a unique solution. Moreover the mapping  $\{\xi^0, s^0, w^0\} \times \{\xi^1, s^1, w^1\} \rightarrow \{\xi, s, w\}$  is continuous from  $\mathcal{V} \times \mathcal{H} \rightarrow C^1([0, T]; \mathcal{V}) \cap C^1([0, T]; \mathcal{H})$ .*

*Proof.* The proof is almost identical to the proof of Theorems 2.1 and 2.2, p. 44 and 47, respectively of [LL]. Thus we mention only the main points. Clearly  $c(\cdot)$  is equivalent to  $\|\cdot\|_{\mathcal{H}}$ . In particular there exists  $\delta > 0$  for which

$$c(\xi, s, w) \geq \delta \|\{\xi, s, w\}\|_{\mathcal{H}}^2.$$

What remains is to obtain a coercive estimate on  $a(\cdot)$  over  $\mathcal{V}$  of the type:

$$(4.4) \quad a(\xi, s, w) + \lambda \|\{\xi, s, w\}\|_{\mathcal{H}}^2 \geq \delta \|\{\xi, s, w\}\|_{\mathcal{V}}^2$$

for some  $\lambda > 0$ ,  $\delta > 0$ . Indeed, if (4.4) holds then by the Lax-Milgram theorem and the usual variational theory (J.L. Lions [Li]) one concludes that (4.1)-(4.3) is well-posed.

In the case  $\Gamma_0 = \emptyset$  (4.4) follows by [La, p.29]. However let us discuss only the case  $\Gamma_0 \neq \emptyset$ . By [LL, p.43]  $a_0(\cdot)$  is coercive over  $H^1(\Omega)$ . Therefore there exists  $\delta_0 > 0$  for which

$$(4.5) \quad a_0(\xi) + 3a_0(s) \geq \delta_0 [\|\xi\|_{(H^1(\Omega))^2}^2 + \|s\|_{(H^1(\Omega))^2}^2].$$



Next, due to Poincaré's inequality, for some  $\delta_1 > 0$

$$(4.6) \quad \|\nabla w\|_{(L^2(\Omega))^2}^2 \geq \delta_1 \|w\|_{H^1(\Omega)}^2.$$

Combining (4.5) and (4.6) it follows that for some  $\delta > 0$

$$a_0(\xi) + 3a_0(s) + a_1(\xi, s, w) > \delta \|\{\xi, s, w\}\|_{\mathcal{V}}.$$

In particular, (4.4) holds with  $\lambda = 0$  when  $\Gamma_0 \neq \emptyset$ .  $\square$

*Remark 4.1.* The system (3.16)-(3.18) reduces exactly to the MTR system in the following two situations:

- (I) **If  $\gamma = 0$**  there is no glue to cause interaction between the top and bottom plate. Hence one would expect to recover the MTR system for a plate of thickness  $h/2$ . Let us see that this is so.

Theorem 4.1, in particular, holds when  $\gamma = 0$ . Thus consider (3.16)-(3.17) in the homogeneous case. Let  $\psi = \xi + 3s$  so that  $\varphi = \psi - \nabla w$ . Adding (ii) and (iii) of (3.16) and dividing the sum by 8 we obtain

$$(4.7) \quad \frac{1}{8} L_\rho \ddot{\psi} + \frac{1}{2} K \varphi - \frac{1}{8} L \psi = 0.$$

Since  $D$  and  $L_\rho$  are cubic in  $h$  while  $K = kEh/2(1 + \mu)$  is linear, if we double  $h$  in (4.7) and (3.16-i) we obtain

$$(4.8) \quad \begin{cases} \text{(i)} & \rho h \ddot{w} + \operatorname{div}(K \varphi) = 0 & \text{in } \Omega \times \mathbb{R}^+ \\ \text{(ii)} & L_\rho \ddot{\psi} + K \varphi - L \psi = 0 & \text{in } \Omega \times \mathbb{R}^+ \\ \text{(iii)} & \varphi = \psi - \nabla w & \text{in } \Omega \times \mathbb{R}^+, \end{cases}$$

In the same way we obtain the boundary conditions

$$(4.9) \quad \begin{cases} \text{(i)} & (w, \psi) = 0 & \text{on } \Gamma_0 \times \mathbb{R}^+ \\ \text{(ii)} & K \varphi \cdot n = 0 & \text{on } \Gamma_1 \times \mathbb{R}^+ \\ \text{(iii)} & \mathcal{B} \psi = 0 & \text{on } \Gamma_1 \times \mathbb{R}^+. \end{cases}$$

Equations (4.8)-(4.9) are precisely the MTR equations for a plate of thickness  $h$ . (We defined  $\psi$  with the opposite sign as in [LL] hence there is a sign difference between [LL, (3.7) p.15] and (4.8)-(4.9).)

- (II) **If  $\gamma \rightarrow \infty$**  the glue becomes infinitely stiff and one expects to see no slip in the limit. Hence one would expect the limiting system to reduce to the MTR system for a plate of thickness  $h$  since the individual plates are themselves MTR plates and by (H1) the shear angles are the same in each layer. In fact one can show that as  $\gamma \rightarrow \infty$  solutions of (3.16)-(3.18) tend in a certain weak sense to solutions of the MTR system (4.8)-(4.9). We make this precise in section 6.

**4.2. Semigroup formulation.** Solutions of (3.16)-(3.18) can also be defined in terms of semigroups on a range of spaces which depend upon the regularities of the body and boundary forces. We limit our discussion here however, for the sake of brevity, to the homogenous case.

First assume  $\Gamma_0 \neq \emptyset$ .

We identify  $\mathcal{H}$  with its dual  $\mathcal{H}'$  and have the dense and continuous embeddings

$$\mathcal{V} \subset \mathcal{H} = \mathcal{H}' \subset \mathcal{V}'.$$

Let  $\langle \cdot, \cdot \rangle$  denote the duality pairing between  $\mathcal{V}$  and  $\mathcal{V}'$  which coincides with  $(\cdot, \cdot)_{\mathcal{H}}$  when both arguments are in  $\mathcal{H}$ . The forms  $c(\cdot; \cdot)$  and  $a(\cdot; \cdot)$  define scalar products on  $\mathcal{H}$  and  $\mathcal{V}$ , respectively, which are equivalent to the natural ones. Thus we may define operators  $C \in \mathcal{L}(\mathcal{H})$  and  $A \in \mathcal{L}(\mathcal{V}, \mathcal{V}')$  by

$$\begin{aligned} \langle C\mathbf{u}, \mathbf{v} \rangle &= c(\mathbf{u}; \mathbf{v}) \quad \forall \mathbf{u}, \mathbf{v} \in \mathcal{H} \\ \langle A\mathbf{u}, \mathbf{v} \rangle &= a(\mathbf{u}; \mathbf{v}) \quad \forall \mathbf{u}, \mathbf{v} \in \mathcal{V}. \end{aligned}$$

Let us set

$$\mathbf{u} = (\xi, s, w) \quad \mathbf{u}^0 = (\xi^0, s^0, w^0) \quad \mathbf{u}^1 = (\xi^1, s^1, w^1)$$

and rewrite (4.2) as

$$(4.10) \quad C\ddot{\mathbf{u}} + A\mathbf{u} = 0 \quad \text{in } \mathcal{V}'.$$

In first order form (4.10) becomes

$$\frac{d}{dt} \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} \dot{\mathbf{u}} \\ \mathbf{u} \end{pmatrix} + \begin{pmatrix} 0 & -A \\ A & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \dot{\mathbf{u}} \end{pmatrix} = 0,$$

or

$$(4.11) \quad C\dot{\mathbf{U}} + \mathcal{A}\mathbf{U} = 0; \quad \mathcal{A} = \begin{pmatrix} 0 & -A \\ A & 0 \end{pmatrix}, \quad C = \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} \mathbf{u} \\ \dot{\mathbf{u}} \end{pmatrix}.$$

Assume we wish to solve (4.11) in  $\mathcal{V} \times \mathcal{H}$ . Define  $\mathcal{D}(A)$  and  $\mathcal{D}(\mathcal{A})$  by

$$\mathcal{D}(A) = \{\mathbf{u} \in \mathcal{H} : A\mathbf{u} \in \mathcal{H}\}, \quad \mathcal{D}(\mathcal{A}) = \mathcal{D}(A) \times \mathcal{V}.$$

Then  $\mathcal{A} : \mathcal{D}(\mathcal{A}) \rightarrow \mathcal{V}' \times \mathcal{H}$  and  $C^{-1} : \mathcal{V}' \times \mathcal{H} \rightarrow \mathcal{V} \times \mathcal{H}$ . Hence we can rewrite (4.11) as

$$(4.12) \quad \dot{\mathbf{U}} + C^{-1}\mathcal{A}\mathbf{U} = 0 \quad \text{in } \mathcal{V} \times \mathcal{H}$$

One can easily verify that  $C^{-1}\mathcal{A}$  is densely defined and furthermore for all  $(\mathbf{u}, \mathbf{v}) \in \mathcal{D}(\mathcal{A})$  we have

$$(4.13) \quad (C^{-1}\mathcal{A}(\mathbf{u}, \mathbf{v}), (\mathbf{u}, \mathbf{v}))_{\mathcal{V} \times \mathcal{H}} = 0.$$

Thus by e.g., Pazy [Pa, p.41],  $-C^{-1}\mathcal{A}$  is the generator of a unitary group on  $\mathcal{V} \times \mathcal{H}$ .

If  $\Gamma_0 = \emptyset$  then  $\mathcal{A}$  is no longer an isomorphism. However  $C$  remains an isomorphism and by (4.4)  $A + \lambda I$  remains an isomorphism  $\mathcal{V} \rightarrow \mathcal{V}'$ . This is enough:

By a simple argument involving the change of variables  $\mathbf{u} = e^{\sigma t} \mathbf{w}$ ,  $\sigma > 0$  it can be shown that  $-\mathcal{C}^{-1} \mathcal{A}$  remains the generator of a  $C_0$ -semigroup when  $\Gamma_0 = \emptyset$ . (See [La; p.32].) Furthermore one can show directly that (4.13) remains valid when  $\Gamma_0 = \emptyset$ . Thus again we reach the conclusion that  $-\mathcal{C}^{-1} \mathcal{A}$  is the generator of a unitary group on  $\mathcal{V} \times \mathcal{H}$ .

By considering appropriate extensions or restrictions of the generator we are lead to the following:

**Theorem 4.2.** *The operator  $\mathcal{C}^{-1} \mathcal{A}$  is the generator of a unitary group on each of the spaces:  $\mathcal{V} \times \mathcal{H}$ ,  $\mathcal{D}(A) \times \mathcal{V}$ ,  $\mathcal{H} \times \mathcal{V}'$ ,  $\mathcal{V}' \times (\mathcal{D}(A))'$ . We thus have the following regularity results for (4.12):*

- (i)  $(\mathbf{u}^0, \mathbf{u}^1) \in \mathcal{D}(A) \times \mathcal{V} \Rightarrow (\mathbf{u}, \dot{\mathbf{u}}) \in C([0, T]; \mathcal{D}(A) \times \mathcal{V})$
- (ii)  $(\mathbf{u}^0, \mathbf{u}^1) \in \mathcal{V} \times \mathcal{H} \Rightarrow (\mathbf{u}, \dot{\mathbf{u}}) \in C([0, T]; \mathcal{V} \times \mathcal{H})$
- (iii)  $(\mathbf{u}^0, \mathbf{u}^1) \in \mathcal{H} \times \mathcal{V}' \Rightarrow (\mathbf{u}, \dot{\mathbf{u}}) \in C([0, T]; \mathcal{H} \times \mathcal{V}')$
- (iv)  $(\mathbf{u}^0, \mathbf{u}^1) \in \mathcal{V}' \times (\mathcal{D}(A))' \Rightarrow (\mathbf{u}, \dot{\mathbf{u}}) \in C([0, T]; \mathcal{V}' \times (\mathcal{D}(A))')$

## 5. Limiting behavior of (3.12)-(3.13) as $K \rightarrow \infty$

When one lets  $K \rightarrow \infty$  in the MTR plate one finds that the solutions converge in a weak sense to solutions of a Kirchhoff plate equation [LL]. In this section we show that as  $K \rightarrow \infty$  solutions of the system (4.1)-(4.3) converge in a weak sense to solutions of a new plate equation which we show is an analog of the Kirchhoff plate.

This limit creates a singular perturbation in which a family of eigenvalues for the shear modes tends to infinity along the imaginary axis. In application,  $K$  is generally several orders of magnitude larger than the other parameters (in particular  $D$ ) which appear in (3.16)-(3.17). Hence in principle, the limiting PDE should provide a good approximation to the low-frequency characteristics of the original system.

Let us recall the variational problem (4.1)-(4.3): Find functions  $\xi = (\xi_1, \xi_2)$ ,  $s = (s_1, s_2)$ ,  $w$  such that

$$(5.1) \quad \begin{cases} (a) & (\xi, s, w) \in C([0, T]; \mathcal{V}) \cap C^1([0, T]; \mathcal{H}), \\ (b) & c(\ddot{\xi}, \ddot{s}, \ddot{w}; \dot{\xi}, \dot{s}, \dot{w}) + a(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) = 0 \quad \forall (\dot{\xi}, \dot{s}, \dot{w}) \in \mathcal{V}, \\ (c) & (\xi(0), s(0), w(0)) = (\xi^0, s^0, w^0) \quad \text{given in } \mathcal{V}, \\ (d) & (\dot{\xi}(0), \dot{s}(0), \dot{w}(0)) = (\xi^1, s^1, w^1) \quad \text{given in } \mathcal{H}. \end{cases}$$

where

$$(5.2) \quad a(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) = a_0(\xi; \dot{\xi}) + 3a_0(s; \dot{s}) + a_1(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) + a_2(s; \dot{s})$$

and  $c, a_0, a_1, a_2$  were defined in (3.4)-(3.8).

### 5.1. The limiting variational problem as $K \rightarrow \infty$ . Let

$$H_{\Gamma_0}^2 = \{w \in H^2(\Omega) \cap H_{\Gamma_0}^1 : \frac{\partial w}{\partial \nu} = 0 \text{ on } \Gamma_0\}$$

and define

$$\begin{aligned} W &= \{(\xi, s, w) \in (H_{\Gamma_0}^1)^4 \times H_{\Gamma_0}^2 : \xi + 3s - \nabla w = 0 \text{ in } \Omega\} \\ V &= \{(\xi, s, w) \in (L^2(\Omega))^4 \times H_{\Gamma_0}^1 : \xi + 3s - \nabla w = 0 \text{ in } \Omega\}. \end{aligned}$$

We will show in Theorem 5.2 that solutions of (5.1) with initial data in  $W \times V$  converge in a weak sense to solutions of the following variational problem:

Find functions  $(\xi, s, w)$  for which

$$(5.3) \quad (\xi, s, w) \in C^1([0, T]; W) \cap C^1([0, T]; V),$$

$$(5.4) \quad \frac{d}{dt} c(\dot{\xi}, \dot{s}, \dot{w}; \dot{\xi}, \dot{s}, \dot{w}) + b(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) = 0 \quad \forall (\dot{\xi}, \dot{s}, \dot{w}) \in W$$

$$(5.5) \quad \begin{cases} (\xi(0), s(0), w(0)) = (\xi^0, s^0, w^0) \text{ given in } W, \\ (\dot{\xi}(0), \dot{s}(0), \dot{w}(0)) = (\xi^1, s^1, w^1) \text{ given in } V, \end{cases}$$

where (5.4) is understood in the sense of distributions on  $(0, T)$  and

$$(5.6) \quad b(\xi, s, w; \dot{\xi}, \dot{s}, \dot{w}) = a_0(\xi; \dot{\xi}) + 3a_0(s; \dot{s}) + a_2(s; \dot{s}),$$

**Theorem 5.1.** *The variational problem (5.3)–(5.5) admits a unique solution. Moreover the mapping  $(\xi^0, s^0, w^0) \times (\xi^1, s^1, w^1) \rightarrow (\xi, s, w)$  is continuous from  $W \times V \rightarrow C^1([0, T]; W) \cap C^1([0, T]; V)$ .*

*Proof.* First assume  $\Gamma_0 \neq \emptyset$ .

We may eliminate  $\xi$  from (5.3)–(5.6) with  $\xi = \nabla w - 3s$ . By (4.5)

$$\begin{aligned} (5.7) \quad a_0(\nabla w - 3s) + 3a_0(s) &\geq \delta_0 \left( \|\nabla w - 3s\|_{(H^1(\Omega))^2}^2 + \|s\|_{H^1(\Omega)^2}^2 \right) \\ &\geq \delta_1 \left( \|\nabla w\|_{(H^1(\Omega))^2}^2 + \|s\|_{(H^1(\Omega))^2}^2 \right), \\ &\geq \delta_2 \left( \|w\|_{H^2(\Omega)}^2 + \|s\|_{(H^1(\Omega))^2}^2 \right), \end{aligned}$$

where the last inequality holds since  $\Gamma_0 \neq \emptyset$ . It follows that  $b(\cdot, \cdot)$  is coercive over  $W$ . Since  $c(\cdot, \cdot)$  is coercive on  $\mathcal{H}$ , we have (by similar inequalities as in (5.7))

$$(5.8) \quad c(\nabla w - 3s, s, w) \geq \delta_3 \left( \|w\|_{H^1(\Omega)}^2 + \|s\|_{L^2(\Omega)}^2 \right).$$

It follows that  $c(\cdot, \cdot)$  is coercive over  $V$  and hence by the standard variational theory (5.3)–(5.5) is well-set. If  $\Gamma_0 = \emptyset$  then (5.7) doesn't hold on  $W$ , however it is enough (see (4.4)) to note that (5.7) holds when  $\lambda((w, w)_\Omega + (s, s)_\Omega)$  are added to the left-hand side. (See [LL, p.46].)  $\square$

**Theorem 5.2.** *Let  $(\xi^0, s^0, w^0), (\xi^1, s^1, w^1)$  be given in  $W \times V$ . Then as  $K \rightarrow \infty$ , the solution  $(\xi_K, s_K, w_K)$  to (5.1) tends to  $(\xi, s, w)$  in the sense that*

$$(5.9) \quad \begin{cases} (\xi_K, s_K, w_K) \rightarrow (\xi, s, w) \text{ in } L^\infty(0, T; \mathcal{V}) \text{ weak star} \\ (\dot{\xi}_K, \dot{s}_K, \dot{w}_K) \rightarrow (\dot{\xi}, \dot{s}, \dot{w}) \text{ in } L^\infty(0, T; \mathcal{H}) \text{ weak star,} \end{cases}$$

where  $\xi, s, w$  are the solutions of (5.3)-(5.5).

*Proof.* This result is similar to [LL, p. 50] where it is shown that the solutions of the Kirchhoff plate equation may be obtained as similar weak limits of solutions of the MTR system. The idea of that proof remains valid here, with the exception that we impose the (additional) compatibility condition  $(\xi^1, s^1, w^1) \in V$  and obtain a stronger result which includes (5.19).

We will assume  $\Gamma_D \neq \emptyset$ . The modifications necessary to handle the general case are the same as those in [LL, p. 51].

Theorem 4.1 shows that (5.1) is well-set and Theorem 4.2 implies the energy

$$(5.10) \quad \mathcal{E}(t) = \frac{1}{2}a(\xi, s, w) + \frac{1}{2}c(\dot{\xi}, \dot{s}, \dot{w})$$

remains constant along trajectories. By (5.5) we have  $a_1(\xi^0, s^0, w^0) = 0$ . Then by conservation of the energy (5.10) as  $K \rightarrow \infty$ ,

$$(5.11) \quad \begin{cases} \{\xi_K, s_K\} & \text{bounded in } L^\infty(0, T; (H_{\Gamma_D}^1)^4) \\ \{\dot{\xi}_K, \dot{s}_K, \dot{w}_K\} & \text{bounded in } L^\infty(0, T; \mathcal{H}) \end{cases}$$

$$(5.12) \quad \{K^{1/2}(\xi_K + 3s_K - \nabla w_K)\} \quad \text{bounded in } L^\infty(0, T; (L^2(\Omega))^2).$$

Therefore, in particular, since  $\Gamma_D \neq \emptyset$

$$(5.13) \quad \{w_K\} \text{ remains bounded in } L^\infty(0, T; H_{\Gamma_D}^1(\Omega)).$$

Thus by extracting a subsequence of  $\{\xi_K, s_K, w_K\}$ , (5.9) follows from (5.11) and (5.13). (Of course, we have yet to identify this limit as a solution of (5.3)-(5.5).) From (5.12) we know  $\{\xi_K + 3s_K - \nabla w_K\} \rightarrow 0$  strongly in  $L^\infty(0, T; (L^2(\Omega))^2)$  and hence also  $\{\dot{\xi}_K + 3\dot{s}_K - \nabla \dot{w}_K\} \rightarrow 0$  in the sense of distributions. Therefore

$$(5.14) \quad \begin{cases} \xi + 3s - \nabla w = 0 & 0 \leq t \leq T \\ \dot{\xi} + 3\dot{s} - \nabla \dot{w} = 0 & \text{almost everywhere } 0 < t < T. \end{cases}$$

Thus from (5.14) and (5.9) one sees that

$$(5.15) \quad (\xi, s, w) \in L^\infty(0, T; W), \quad (\dot{\xi}, \dot{s}, \dot{w}) \in L^\infty(0, T; V).$$

To see that  $(\xi, s, w)$  satisfy (5.4), we take in (5.1) test functions  $(\hat{\xi}, \hat{s}, \hat{w}) \in W$  so that (5.1-b) (with  $\xi = \xi_K, s = s_K, w = w_K$ ) reduces to

$$(5.16) \quad \frac{d}{dt}c(\dot{\xi}_K, \dot{s}_K, \dot{w}_K; \dot{\hat{\xi}}, \dot{\hat{s}}, \dot{\hat{w}}) + a_0(\xi_K; \hat{\xi}) + 3a_0(s_K; \hat{s}) + a_2(s_K; \hat{s}) = 0.$$

Using (5.9) we pass to the limit (in the sense of distributions) and obtain (5.4).

Let us now check that  $(\xi(0), s(0), w(0)) = (\xi^0, s^0, w^0)$ . By (5.11), (5.13) we know that  $\{(\xi_K, s_K, w_K)\}$  is bounded in  $L^\infty(0, T; \mathcal{V}) \cap W^{1, \infty}(0, T; \mathcal{H})$ , and thus by compactness, in addition to (5.9) we have

$$(\xi_K, s_K, w_K) \rightarrow (\xi, s, w) \quad \text{in } C([0, T]; \mathcal{H}).$$

Consequently by passing limits we obtain

$$(5.17) \quad (\xi(0), s(0), w(0)) = (\xi^0, s^0, w^0).$$

As is well-known, the solution to (5.1) is the same as the semigroup solution in (ii) of Theorem 4.2. When this solution is differentiated with respect to time it is easy to see from (4.12) that  $(\dot{\mathbf{u}}, \dot{\mathbf{u}}) \in C([0, T]; \mathcal{H} \times \mathcal{V})$ . Thus for each  $K$   $(\dot{\xi}_K, \dot{s}_K, \dot{w}_K) \in C([0, T]; \mathcal{V}')$  and if we use the hypothesis that  $(\xi^0, s^0, w^0) \in W$ ,  $(\xi^1, s^1, w^1) \in V$ , then (4.12) implies  $\dot{w}_K(0) = 0$  and  $\{\dot{s}_K(0), \dot{\xi}_K(0)\} \in ((H_{\Gamma_0}^1)')^2$  independent of  $K$ . Furthermore by Theorem 4.2 we also know that conservation of (norm) energy holds in that space. Thus we are in a position to repeat with  $\{\dot{\xi}_K, \dot{s}_K, \dot{w}_K\}$  the same argument that was applied to  $\{\xi_K, s_K, w_K\}$ . We obtain

$$(5.18) \quad \{\dot{\xi}_K, \dot{s}_K, \dot{w}_K\} \quad \text{remain bounded in } L^\infty(0, T; \mathcal{V}').$$

Consequently

$$\{\dot{\xi}_K, \dot{s}_K, \dot{w}_K\} \rightarrow (\dot{\xi}, \dot{s}, \dot{w}) \quad \text{in } C([0, T]; (\mathcal{V}')),$$

and hence

$$(5.19) \quad (\dot{\xi}(0), \dot{s}(0), \dot{w}(0)) = (\xi^1, s^1, w^1).$$

Since the initial data were given in  $W \times V$ , by Theorem 5.1 there is a unique solution  $(\tilde{\xi}, \tilde{s}, \tilde{w})$  to (5.4)-(5.5) within the space  $C([0, T]; W) \cap C^1([0, T]; V)$ . Furthermore by a regularity improvement theorem in Lions and Magenes [LM, p.275], (5.15) implies that after modification on a set of measure zero,

$$(5.20) \quad (\xi, s, w) \in C([0, T]; W) \cap C^1([0, T]; V).$$

Hence  $(\xi, s, w) = (\tilde{\xi}, \tilde{s}, \tilde{w})$  is uniquely determined and satisfies (5.3)-(5.5).  $\square$

*Remark 5.1.* If the hypothesis that  $(\xi^1, s^1, w^1)$  is weakened to

$$(\xi^1, s^1, w^1) \in \mathcal{H} : w^1 \in H_{\Gamma_0}^1$$

then (5.18) will not in general hold and consequently one can not infer (5.19). However as a consequence of (5.9) and (5.15) there will exist limit points satisfying (5.17), (5.20).

**5.2. Semigroup formulation of limiting system.** The semigroup formulation of (5.3)-(5.5) is analogous to that of (5.1) and a result similar to Theorem 4.2 also holds here. Thus for simplicity we assume for this discussion that  $\Gamma_0 \neq \emptyset$ .

We introduce the forms

$$(5.21) \quad c_1(s, w; \hat{s}, \hat{w}) = c(\nabla w - 3s, s, w; \nabla \hat{w} - 3\hat{s}, \hat{s}, \hat{w})$$

$$(5.22) \quad b_0(s, w; \hat{s}, \hat{w}) = a_0(\nabla w - 3s; \nabla \hat{w} - 3\hat{s}) + 3a_0(s; \hat{s}),$$

and continue to use the notation  $c_1(s, w) = c_1(s, w; s, w)$  and likewise for the form  $b_0$ .

When  $\xi$  and  $\hat{\xi}$  are eliminated from (5.3)-(5.5) the variational equation may be written

$$(5.23) \quad \frac{d}{dt} [c_1(\dot{s}, \dot{w}; \hat{s}, \hat{w})] + b_0(s, w; \hat{s}, \hat{w}) + a_2(s, \hat{s}) = 0, \quad \forall (\dot{w}, \hat{s}) \in H_{\Gamma_0}^2 \times (H_{\Gamma_0}^1)^2,$$

with initial conditions

$$(5.24) \quad \begin{cases} \{s(0), w(0)\} = \{s^0, w^0\} & \text{given in } W_0 \cong (H_{\Gamma_0}^1)^2 \times H_{\Gamma_0}^2 \\ \{\dot{s}(0), \dot{w}(0)\} = \{\dot{s}^1, \dot{w}^1\} & \text{given in } V_0 = (L^2(\Omega))^2 \times H_{\Gamma_0}^1. \end{cases}$$

It is easy to see that  $c_1(\cdot; \cdot)$  defines a scalar product on  $V_0$  equivalent to the natural one and since  $\Gamma_0 \neq \emptyset$ ,  $b_0(\cdot; \cdot)$  defines a scalar product on  $W_0$  equivalent to the natural one. Let  $\langle \cdot; \cdot \rangle$  defined on  $W_0' \times W_0$  be the duality pairing of  $W_0'$  and  $W_0$  with respect to  $H' = H = (L^2(\Omega))^3$ .

We thus have the continuous dense embeddings

$$W_0 \subset V_0 \subset H \subset V_0' \subset W_0'$$

Define  $A \in \mathcal{L}(W_0, W_0')$  and  $C \in \mathcal{L}(V_0, V_0')$  by

$$\begin{aligned} \langle A\{s, w\}, \{\hat{s}, \hat{w}\} \rangle &= b_0(s, w; \hat{s}, \hat{w}) + a_2(s; \hat{s}), \\ \langle C\{\dot{s}, \dot{w}\}, \{\hat{s}, \hat{w}\} \rangle &= c_1(s, w; \hat{s}, \hat{w}). \end{aligned}$$

Following the approach of section 4.2 we write (5.23) as

$$(5.25) \quad \dot{\mathbf{U}} + \mathcal{C}^{-1} \mathcal{A} \mathbf{U} = 0; \quad \mathcal{A} = \begin{pmatrix} 0 & -A \\ A & 0 \end{pmatrix}, \quad \mathcal{C} = \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix},$$

where  $\mathbf{U} = \{\mathbf{u}, \mathbf{v}\}^T$ ;  $\mathbf{u} = (s, w)$ ,  $\mathbf{v} = (\dot{s}, \dot{w})$ . (Here  $T$  denotes transposition.) The domains of  $A$  and  $\mathcal{A}$  are defined by

$$\mathcal{D}(A) = \{\mathbf{u} \in W_0 : A\mathbf{u} \in V_0'\}, \quad \mathcal{D}(\mathcal{A}) = \{\{\mathbf{u}, \mathbf{v}\} : \mathbf{u} \in \mathcal{D}(A), \mathbf{v} \in W_0\}.$$

Then just as in section 4.2,  $\mathcal{C}^{-1} \mathcal{A} : \mathcal{D}(\mathcal{A}) \rightarrow W_0 \times V_0$  is densely defined and also satisfies (4.13). Thus by a proof similar to the one Theorems 4.2 we have the following.

**Theorem 5.3.** *The operator  $-\mathcal{C}^{-1}\mathcal{A}$  is the generator of a unitary group on any of the spaces  $\mathcal{D}(A) \times W_0$ ,  $W_0 \times V_0$ ,  $V'_0 \times W'_0$ ,  $W'_0 \times (\mathcal{D}(A))'$ . Furthermore one has the following regularity results*

- (i)  $(\mathbf{u}^0, \mathbf{u}^1) \in \mathcal{D}(A) \times W_0 \Rightarrow (\mathbf{u}, \dot{\mathbf{u}}) \in C([0, T]; \mathcal{D}(A) \times W_0)$
- (ii)  $(\mathbf{u}^0, \mathbf{u}^1) \in W_0 \times V_0 \Rightarrow (\mathbf{u}, \dot{\mathbf{u}}) \in C([0, T]; W_0 \times V_0)$
- (iii)  $(\mathbf{u}^0, \mathbf{u}^1) \in V'_0 \times W'_0 \Rightarrow (\mathbf{u}, \dot{\mathbf{u}}) \in C([0, T]; V'_0 \times W'_0)$
- (iv)  $(\mathbf{u}^0, \mathbf{u}^1) \in W'_0 \times (\mathcal{D}(A))' \Rightarrow (\mathbf{u}, \dot{\mathbf{u}}) \in C([0, T]; W'_0 \times (\mathcal{D}(A))')$ .

**5.3. Boundary value problem for limiting system as  $K \rightarrow \infty$ .** Assume the plate is subject to body forces  $(f_3, M, \mathcal{K})$  and boundary forces  $(g_3, m, \kappa)$  as in (3.10), (3.16), (3.17). (More precisely,  $M, m, \mathcal{K}, \kappa$  are force moments.) For consistency with no shear we impose

$$(5.26) \quad \int_{-h/2}^0 \tilde{f}_i(x_3 + h/4) dx_3 = \int_0^{h/2} f_i(x_3 - h/4) dx_3 = 0 \quad i = 1, 2.$$

Adding the two equations in (5.26) gives  $M_i(\underline{x}) - \frac{h}{4} P_i(\underline{x}) = 0 \quad \forall \underline{x} \in \Omega$ . Consequently from (3.9) we find  $M(\underline{x}) = \mathcal{K}(\underline{x}) \quad \forall \underline{x} \in \Omega$ . Likewise from (3.10) the same condition holds on  $\Gamma_1$ . Thus for compatibility with lack of shear notions, we impose the conditions

$$(5.27) \quad \begin{cases} M_i = \mathcal{K}_i & i = 1, 2 & \text{in } \Omega \times (0, T) \\ m_i = \kappa_i & i = 1, 2 & \text{on } \Gamma_1 \times (0, T) \end{cases}$$

When the force terms are included with (4.2) (from (3.2)) we have

$$0 = c(\ddot{\xi}, \ddot{s}, \ddot{w}; \dot{\xi}, \dot{s}, \dot{w}) + a_0(\xi; \dot{\xi}) + 3a_0(s; \dot{s}) + a_2(s; \dot{s}) \\ - \int_{\Omega} f_3 \dot{w} - M \cdot \dot{\xi} + \mathcal{K} \cdot \dot{s} d\underline{x} - \int_{\Gamma_1} g_3 \dot{w} - m \cdot \dot{\xi} + \kappa \cdot \dot{s} d\Gamma.$$

We then use  $\hat{\xi} = \nabla \dot{w} - 3\dot{s}$ , and (5.27) to obtain

$$0 = c(\ddot{\xi}, \ddot{s}, \ddot{w}; \nabla \dot{w} - 3\dot{s}, \dot{s}, \dot{w}) + a_0(\xi; \nabla \dot{w} - 3\dot{s}) + 3a_0(s; \dot{s}) + a_2(s; \dot{s}) \\ - \int_{\Omega} f_3 \dot{w} - M \cdot \nabla \dot{w} + 4M \cdot \dot{s} d\underline{x} - \int_{\Gamma_1} g_3 \dot{w} - m \cdot \nabla \dot{w} + 4m \cdot \dot{s} d\Gamma.$$

By Green's theorem one finds



$$\begin{aligned}
0 = & \int_{\Omega} \hat{w}(\rho h \ddot{w} - \operatorname{div}(I_{\rho}(\ddot{\xi}) - L\xi)) + 3\hat{s} \cdot (I_{\rho}\ddot{s} - Ls + \frac{4}{3}h^2\gamma s - I_{\rho}(\ddot{\xi}) + L\xi) d\underline{x} \\
& + \int_{\Gamma_1} (\nabla \hat{w} - 3\hat{s}) \cdot \mathcal{B}\xi + 3\hat{s} \cdot \mathcal{B}(s) + \hat{w} \left[ I_{\rho} \left( \frac{\partial \ddot{w}}{\partial n} - 3\ddot{s} \cdot n \right) - (L\xi) \cdot n \right] d\Gamma \\
& - \int_{\Omega} f_3 \hat{w} + (\operatorname{div} M) \hat{w} + 4M \cdot \hat{s} d\underline{x} - \int_{\Gamma_1} g_3 - M \cdot n \hat{w} - m \cdot \nabla \hat{w} + 4m \cdot \hat{s} d\Gamma.
\end{aligned}$$

Either  $\Gamma_0 = \emptyset$  or  $w$  vanishes on  $\Gamma_0$ . Either way we may use the identity

$$(5.28) \quad \int_{\Gamma_1} \nabla \hat{w} \cdot \psi d\Gamma = \int_{\Gamma_1} \frac{\partial \hat{w}}{\partial n} \psi \cdot n - \hat{w} \frac{\partial}{\partial \tau} (\psi \cdot \tau) d\Gamma$$

where  $n = (n_1, n_2)$  is the unit normal vector,  $\tau = (-n_2, n_1)$  is the unit tangent vector to  $\Gamma$  and  $\psi = (\psi_1, \psi_2)$ . By collecting terms and applying (5.28) we obtain the boundary value problem:

$$(5.29) \quad \begin{cases} \rho h \ddot{w} - \operatorname{div}(I_{\rho} \ddot{\xi}) + \operatorname{div} L(\xi) = f_3 + \operatorname{div} M \\ 4I_{\rho} \ddot{s} - I_{\rho} \nabla \ddot{w} + L(\xi - s) + \frac{4}{3}h^2\gamma s = \frac{4}{3}M \\ \xi = \nabla w - 3s \end{cases} \quad \text{in } \Omega \times (0, T)$$

$$(5.30) \quad w = \frac{\partial w}{\partial n} = s = 0 \quad \text{on } \Gamma_0 \times (0, T)$$

$$(5.31) \quad \begin{cases} I_{\rho} \left( \frac{\partial \ddot{w}}{\partial n} - 3\ddot{s} \cdot n \right) + J_1 \xi = g_3 - M \cdot n + \frac{\partial}{\partial \tau} (m \cdot \tau) \\ J_2 \xi = -m \cdot n \\ \mathcal{B}(4s - \nabla w) = \frac{4}{3}m \end{cases} \quad \text{on } \Gamma_1 \times (0, T)$$

where  $J_1 \psi = -(L\psi) \cdot n - \frac{\partial}{\partial \tau} (\mathcal{B}\psi \cdot \tau)$  and  $J_2 \psi = \mathcal{B}\psi \cdot n$ . Initial conditions may be given as in (5.5).

*Remark 5.2.* One sees that the boundary value problem (5.29)-(5.31) is an analog of the Kirchhoff plate equations [LL, p. 13]. Indeed, in the following two situations, the system (5.29)-(5.31) reduces exactly to a Kirchhoff system.

- (I) **If  $\gamma = 0$ ,** then there is no interaction between the two plates. Thus one expects to obtain a Kirchhoff plate with thickness  $h/2$ . For simplicity, assume there are no external forces and  $\Gamma_0 \neq \emptyset$ . Furthermore assume the initial conditions satisfy  $(s^0, s^1) = (\nabla w^0/4, \nabla w^1/4)$ . We know by Theorem 5.1 that the system is well posed when  $\gamma = 0$ . Let  $y = 4s - \nabla w$ . Then from (5.29),

$$(5.32) \quad I_{\rho} \ddot{y} - Ly = 0 \quad \text{in } \Omega \times (0, T)$$

with boundary conditions

$$(5.33) \quad y = 0 \text{ on } \Gamma_0 \times (0, T), \quad \mathcal{B}y = 0 \text{ on } \Gamma_1 \times (0, T).$$

Since  $-L$  is a positive second order operator (corresponding to the positive quadratic form  $a_0$ ) and  $\mathcal{B}$  is the associated conormal derivative operator (see (3.15)) there exists a unique solution to (5.32)-(5.33) with initial conditions

$$y(0) = \dot{y}(0) = 0.$$

Thus  $y = 0$  in  $\Omega \times (0, T)$  and we obtain

$$s = \frac{\nabla w}{4} \text{ in } \Omega \times (0, T).$$

What remains is

$$(5.34) \quad \begin{cases} \rho \frac{h}{2} \ddot{w} - \operatorname{div} \left( \frac{I_p}{\kappa} \nabla \ddot{w} \right) + \frac{1}{\kappa} \operatorname{div} L(\nabla w) = 0 & \text{in } \Omega \times (0, T) \\ w = \frac{\partial w}{\partial n} = 0 & \text{on } \Gamma_0 \times (0, T) \\ \frac{I_p}{\kappa} \left( \frac{\partial \ddot{w}}{\partial n} \right) + \frac{1}{\kappa} J_1 \nabla w = 0 & \text{on } \Gamma_1 \times (0, T) \\ \frac{1}{\kappa} J_2 \nabla w = 0 & \text{on } \Gamma_1 \times (0, T) \\ (w(0), \dot{w}(0)) & \text{given in } H_{\Gamma_0}^2 \times H_{\Gamma_0}^1. \end{cases}$$

One easily sees that (5.34) is precisely the Kirchhoff system for a plate of thickness  $h/2$ . (See Remark 4.1 for the effect of the scaling.)

- (II) If  $\gamma \rightarrow \infty$ , then one expects the system (5.29)-(5.31) to reduce to a Kirchhoff system for a plate of thickness  $h$  since infinitely strong glue will prevent any slip from occurring. As  $\gamma \rightarrow \infty$  we expect  $s \rightarrow 0$ . Putting  $s \equiv 0$  in (5.29)-(5.31) gives precisely the Kirchhoff system for a plate of thickness  $h$ :

$$(5.35) \quad \begin{cases} \rho h \ddot{w} - \operatorname{div} (I_p \nabla \ddot{w}) + \operatorname{div} L(\nabla w) = 0 & \text{in } \Omega \times (0, T) \\ w = \frac{\partial w}{\partial n} = 0 & \text{on } \Gamma_0 \times (0, T) \\ I_p \left( \frac{\partial \ddot{w}}{\partial n} \right) + J_1 \nabla w = 0 & \text{on } \Gamma_1 \times (0, T) \\ J_2 \nabla w = 0 & \text{on } \Gamma_1 \times (0, T) \\ (w(0), \dot{w}(0)) & \text{given in } H_{\Gamma_0}^2 \times H_{\Gamma_0}^1. \end{cases}$$

We discuss this situation in more detail in the next section.

## 6. Behavior of (5.1) and (5.3)-(5.5) as $\gamma \rightarrow \infty$

If  $\gamma \rightarrow \infty$  then one expects the system (5.1) to reduce to the MTR system for a plate of thickness  $h$  since in the limit there will be no slip. Likewise, since one obtains the Kirchhoff system as a weak limit as  $K \rightarrow \infty$  of solutions of the MTR system [LL], one would expect to obtain a Kirchhoff plate of thickness  $h$  as a weak limit as  $\gamma \rightarrow \infty$  of solutions of (5.3)-(5.5). We make precise statements to this effect in this section.

**6.1. Behavior of (5.1) as  $\gamma \rightarrow \infty$ .** The MTR system (4.8)-(4.9) may be written in variational form as

$$(6.1) \quad \frac{d}{dt} c(\dot{\psi}, 0, \dot{w}; \hat{\psi}, 0, \hat{w}) + a_0(\psi; \hat{\psi}) + a_1(\psi, 0, w; \hat{\psi}, 0, \hat{w}) = 0, \quad \forall (\hat{\psi}, \hat{w}) \in (H_{\Gamma_0}^1)^3.$$

For each initial data  $\{(\psi^0, w^0), (\psi^1, w^1)\} \in (H_{\Gamma_0}^1)^3 \times (L^2(\Omega))^3$  there is a unique solution (see [LL]) with

$$(6.2) \quad (\psi, w) \in C([0, T]; (H_{\Gamma_0}^1)^3) \cap C^1([0, T]; (L^2(\Omega))^3).$$

**Theorem 6.1.** *Let  $(\psi^0, s^0, w^0), (\psi^1, s^1, w^1)$  be given in  $\mathcal{H} \times \mathcal{V}$  with  $s^0 = s^1 = 0$ . Then as  $\gamma \rightarrow \infty$ , the solution  $(\psi_\gamma, s_\gamma, w_\gamma)$  to (5.1) tends to  $(\psi, 0, w)$  in the sense that  $s \rightarrow 0$  in  $L^\infty(0, T; L^2(\Omega))$  and*

$$(6.3) \quad \begin{cases} (\psi_\gamma, w_\gamma) \rightarrow (\psi, w) & \text{in } L^\infty(0, T; (H_{\Gamma_0}^1)^3) \text{ weak star} \\ (\dot{\psi}_\gamma, \dot{w}_\gamma) \rightarrow (\dot{\psi}, \dot{w}) & \text{in } L^\infty(0, T; (L^2(\Omega))^3) \text{ weak star.} \end{cases}$$

where  $(\psi, w)$  are the solutions of (6.1).

*Sketch of Proof.* Since the proof involves the same steps as the proof of Theorem 5.2 we only mention the main points. For simplicity we assume  $\Gamma_0 \neq \emptyset$ .

We use conservation of the energy (5.10) and find that

$$(6.4) \quad \begin{cases} \{\psi_\gamma, s_\gamma\} & \text{bounded in } L^\infty(0, T; (H_{\Gamma_0}^1)^4) \\ \{\dot{\psi}_\gamma, \dot{s}_\gamma, \dot{w}_\gamma\} & \text{bounded in } L^\infty(0, T; \mathcal{H}) \end{cases}$$

$$(6.5) \quad \{\psi_\gamma + 3s_\gamma - \nabla w_\gamma\} \quad \text{bounded in } L^\infty(0, T; (L^2(\Omega))^2)$$

$$(6.6) \quad \{\gamma^{1/2}s\} \quad \text{bounded in } L^\infty(0, T; (L^2(\Omega))^2).$$

Thus  $s \rightarrow 0$  strongly as  $\gamma \rightarrow \infty$ . Since  $\Gamma_0 \neq \emptyset$  (6.5) implies that  $w_\gamma$  remains bounded in  $L^\infty(0, T; H_{\Gamma_0}^1)$ . Thus by taking a subsequence (6.3) holds.

It remains to show that  $(\psi, w)$  are defined by (6.1)-(6.2). We take in (5.2) test functions  $\hat{\psi}, \hat{s}, \hat{w}$  with  $\hat{s} = 0$  so (5.2) (with  $\psi = \psi_\gamma, s = s_\gamma, w = w_\gamma$ ) reduces to

$$\frac{d}{dt} c(\dot{\psi}_\gamma, 0, \dot{w}_\gamma; \hat{\psi}, 0, \hat{w}) + a_0(\psi_\gamma; \hat{\psi}) + a_1(\psi_\gamma, 0, w_\gamma; \hat{\psi}, 0, \hat{w}) = 0 \quad \forall (\hat{\psi}, \hat{w}) \in (H_{\Gamma_0}^1)^3.$$

Using (6.3) we pass to the limit and obtain (6.1). Finally we need to check that  $(\psi(0), w(0)) = (\psi^0, w^0)$  and  $(\dot{\psi}(0), \dot{w}(0)) = (\psi^1, w^1)$ . The first of these follows from the boundedness of the time derivatives in (6.4). To obtain the second we differentiate (4.12) and note that by Theorem 4.2 the resulting equation in the variables  $(\dot{\mathbf{u}}_\gamma, \dot{\mathbf{u}}_\gamma)$  is well-posed on  $\mathcal{H} \times \mathcal{V}'$  and the norm of the solution at time  $t$  is conserved. Since  $s^1 = 0$  we may repeat the same argument to obtain

$$\{\dot{\psi}_\gamma, \dot{s}_\gamma, \dot{w}_\gamma\} \quad \text{remain bounded in } L^\infty(0, T; \mathcal{V}').$$

Thus we obtain convergence of the initial values in  $\mathcal{H} \times \mathcal{V}'$ . However since the data was given in  $\mathcal{V} \times \mathcal{H}$  and the MTR system is well-posed on that space we actually have the regularity in (6.2).  $\square$

**6.2. Behavior of (5.3)-(5.5) as  $\gamma \rightarrow \infty$ .** The Kirchhoff system (5.36) can be written in variational form as

$$(6.7) \quad \frac{d}{dt} [c(\nabla \dot{w}, 0, \dot{w}; \nabla \dot{w}, 0, \dot{w})] + a_0(\nabla w, \nabla \dot{w}) = 0 \quad \forall \dot{w} \in H_{\Gamma_0}^2$$

$$(6.8) \quad w \in C^1([0, T]; H_{\Gamma_0}^2) \cap C^1([0, T]; H_{\Gamma_0}^1)$$

$$(6.9) \quad (w(0), \dot{w}(0)) = (w^0, w^1) \quad \text{in } H_{\Gamma_0}^2 \times H_{\Gamma_0}^1.$$

By [LL, p.48], (6.7)-(6.9) has a unique solution.

The variational problem (5.3)-(5.5) was shown to have a unique solution in Theorem 5.1. We can eliminate the variable  $\xi$  and rewrite (5.3)-(5.5) in the form (5.23)-(5.24) and in this form, the problem has a unique solution with

$$(6.10) \quad (s, w) \in C^1([0, T]; W_0) \cap C^1([0, T]; V_0),$$

where the spaces  $W_0$  and  $V_0$  are defined by (5.24).

**Theorem 6.2.** *Let  $(s^0, w^0), (s^1, w^1)$  be given in  $W_0 \times V_0$  with  $s^0 = s^1 = 0$ . Then as  $\gamma \rightarrow \infty$ , the solution  $(s_\gamma, w_\gamma)$  to (5.23)-(5.24) tends to  $(0, w)$  in the sense that  $s_\gamma \rightarrow 0$  in  $L^\infty(0, T; L^2(\Omega))$  and*

$$(6.11) \quad \begin{cases} \{w_\gamma\} \rightarrow w & \text{in } L^\infty(0, T; H_{\Gamma_0}^2) \text{ weak star} \\ \{\dot{w}_\gamma\} \rightarrow \dot{w} & \text{in } L^\infty(0, T; H_{\Gamma_0}^1) \text{ weak star.} \end{cases}$$

where  $w$  is the solution to (6.7)-(6.9).

*Sketch of Proof.* Again, let us mention only the main points. The details can be filled in following the proof of Theorem 5.2. For simplicity assume  $\Gamma_0 \neq \emptyset$ .

From Theorem 5.3 the energy

$$\mathcal{E}(t) = \frac{1}{2}c_1(s, w) + \frac{1}{2}b_0(s, w) + \frac{1}{2}a_2(s)$$

is conserved along all solution trajectories of (5.23)-(5.24). From this we know

$$(6.12) \quad \{\nabla w_\gamma - 3s_\gamma, s_\gamma\} \quad \text{bounded in } L^\infty(0, T; (H_{\Gamma_0}^1)^4)$$

$$(6.13) \quad \{\nabla \dot{w}_\gamma - 3\dot{s}_\gamma, \dot{s}_\gamma, \dot{w}_\gamma\} \quad \text{bounded in } L^\infty(0, T; (L^2(\Omega))^5)$$

$$(6.14) \quad \{\gamma^{1/2}s\} \quad \text{bounded in } L^\infty(0, T; (L^2(\Omega))^2).$$

Thus  $s \rightarrow 0$  strongly in  $L^\infty(0, T; L^2(\Omega)^2)$  as  $\gamma \rightarrow \infty$ . Since  $\Gamma_0 \neq \emptyset$  (6.12) implies that  $w_\gamma$  remains bounded in  $L^\infty(0, T; H_{\Gamma_0}^2)$ . Thus by taking a subsequence (6.11) holds.

To see that  $w$  satisfies (6.7) we take in (5.23) test functions  $\hat{s}, \hat{w}$  with  $\hat{s} = 0$  so (5.23) (with  $s = s_\gamma$ ,  $w = w_\gamma$ ) reduces to

$$(6.15) \quad \frac{d}{dt} c(\dot{\nabla} w_\gamma - 3s_\gamma, 0, \dot{w}_\gamma; \nabla \hat{w}, 0, \hat{w}) + a_0(\nabla w_\gamma - 3s_\gamma; \nabla \hat{w}) = 0 \quad \forall \hat{w} \in H_{\Gamma_0}^2.$$

Since  $s_\gamma \rightarrow 0$  strongly in  $L^2(\Omega)$  and by (6.12)  $s_\gamma$  converges weakly to a limit function in  $L^\infty(0, T; H_{\Gamma_0}^1)$  this limit must also be zero. Thus passing the limit in (6.15) gives (6.7). What remains is to obtain the convergence of the initial values (6.9). This is done exactly in the same way as in the proof of Theorem 5.2, i.e., we use Theorem 5.3 to infer essential boundedness of  $\dot{w}$  in a weaker space and use this to infer (6.9), but in the weaker space. Since the initial data were given in  $H_{\Gamma_0}^2 \times H_{\Gamma_0}^1$  we then by well posedness of (6.7)-(6.9) obtain the regularity in (6.9).  $\square$

**Acknowledgements.** The author is indebted to J.E. Lagnese for many very helpful suggestions and also to Jiongmin Yong and Enrique Zuazua for their help on some technical points. Most of this research was conducted while the author was supported by Institute for Mathematics and its Applications (IMA) of the University of Minnesota during the "Year in Control Theory" 1992-93. The author is grateful for their hospitality.

## References

- [Di] DiTaranto, R.A.: *Theory of vibratory bending for elastic and viscoelastic layered finite-length beams*, J. Appl. Mech. **32** (1965), 881-886.
- [He] Heuer, R.: *Static and dynamic analysis of transversely isotropic, moderately thick sandwich beams by analogy*, Acta Mechanica **91** (1992), 1-9.
- [Ke] Kerwin, E.M.: *Damping of flexural waves by constrained visco-elastic layer*, J. Acoustical Soc. of Am. **31** (1959), 952-962.
- [La] Lagnese, J.E.: *Boundary Stabilization of Thin Plates*, series "SIAM Studies in Applied Mathematics" SIAM, Philadelphia 1989.
- [LL] Lagnese, J.E. and Lions, J.-L.: *Modelling Analysis and Control of Thin Plates*, collection: "Recherches en Mathématiques Appliquées", RMA 6, Springer-Verlag, New York, 1989.
- [Li] Lions, J.-L.: *Equations Différentielles Opérationnelles et Problèmes aux Limites*, Grundlehren B III, Springer-Verlag, Berlin 1961.
- [LM] Lions, J.-L. and Magenes, E.: *Non-Homogeneous Boundary Value Problems and Applications*, Vol 1, Springer-Verlag, Berlin 1972.
- [Me] Mead, D.J.: *A comparison of some equations for the flexural vibration of damped sandwich beams*, J. Sound Vib. **83** (1982), 363-377.
- [Pa] Pazy, A.: *Semigroups of linear operators and applications to partial differential equations*, Springer-Verlag series "Applied Mathematical Sciences" # 44 Springer-Verlag, New York, 1983.
- [Ti] Timoshenko, S., Young, D.H., Weaver, W.: *Vibration problems in engineering*, New York, Wiley, 1974.
- [YD] Yun, M.-J., Dowell, E.H.: *Governing equations for vibrating constrained-layer damping sandwich plates and beams*, J. Appl. Mech. **39** (1972) 1041-1046.
- [Yu] Yu, Y.-Y.: *Simple thickness-shear modes of vibration of infinite sandwich plates*, J. Appl. Mech. **26** (1959), 679-681.

Scott W. Hansen  
Department of Mathematics  
Iowa State University,  
Ames, Iowa 50011, USA

## NUMERICAL SOLUTION OF A CONSTRAINED CONTROL PROBLEM FOR A PHASE FIELD MODEL

M. HEINKENSCHLOSS AND E.W. SACHS\*

Interdisciplinary Center for Applied Mathematics  
Department of Mathematics  
Virginia Polytechnic Institute and State University

FB IV Mathematik  
Universität Trier

**ABSTRACT.** A phase field model is considered for an evolution process which describes the phase change from solid to liquid. The process is controlled by inputs which are uniformly bounded. The objective is that the phase and the temperature follow certain desired functions as close as possible with a cost term for the control. The necessary optimality conditions of first order are formulated for this control problem. These conditions include projections which make the application of Newton's method difficult. We present an approach to circumvent this difficulty and obtain an algorithm which is very efficient numerically. The numerical examples indicate that the required time to solve the problem with control constraints is of the same magnitude as for the unconstrained problem.

**1991 Mathematics Subject Classification.** 49M15

**Key words and phrases.** Optimal control, bound constraints, projected Newton method, phase field model.

### 1. Introduction

Phase field models describe the solid-liquid phase transition in a pure material. They involve the temperature of the material and the phase function which indicates the liquid or solid state of the material. These models consist of a system of two nonlinear parabolic differential equations, [4], [7], [8], [13], [15]. We consider the phase field model in [4], [7], [8].

The corresponding control problem is to follow a prescribed temperature and/or phase by controlling the temperature through a system of diffusion equations. Problems of this type have been considered in [5], [6], [11]. For these problems

---

\*This work was supported in part by AFOSR under Grant F49620-93-1-0280 while the author was a visiting scientist at the Interdisciplinary Center for Applied Mathematics, Virginia Polytechnic Institute and State University.

necessary optimality conditions were obtained. Computations in [6] were done using a gradient method combined with a discretization scheme of the differential equation. In [10] a different approach was presented which exploits the necessary optimality conditions, a system of four coupled nonlinear partial differential equations. The problem was reformulated and Newton's method was applied. Due to the reformulation a decoupled system needed to be solved at each step. This was carried out for a discretized version using GMRES with a hierarchical basis preconditioner.

In this paper we extend the model and add simple bounds as constraints on the controls. Existence of optimal controls and necessary optimality conditions have been shown in [11]. We rewrite these and obtain a fixed point problem for the optimal control, temperature, phase, and their adjoints. A classical approach for the solution of these box constrained problems is the gradient projection method. However, this method is a first order method and converges rather slowly. Moreover, it requires the solution of the nonlinear state equation in each step. Newton's method, on the other hand, cannot be applied directly to the optimality system because the projection involved for the controls is nondifferentiable.

It is the goal of this paper to obtain a projected Newton algorithm with favorable convergence properties. For minimization problems in finite dimensions the gradient projection method has been adapted in [2] to successfully utilize Newton steps. In [12] an extension of this approach is used to solve the necessary optimality conditions in infinite dimensions obtained from a control problem with ordinary differential equations and control constraints. The advantage over the approach in [2] is that at each step only the linearized state equation needs to be solved instead of the original nonlinear one. A convergence proof in infinite dimensions is given for an ordinary control problem with a control in one dimension.

For the phase field control problem we can also formulate an algorithm which leads to a projected Newton method. To do this we extend the approach in [12] to control problems governed by systems of parabolic equations and combine this with the multilevel Newton method used in [10] for the unconstrained problem. Each step requires the solution of a system of linear boundary value problems where the decoupling strategy [10] from the unconstrained case can be used again. This yields an efficient solver for the linearized state and adjoint equations. As in the unconstrained case this allows to eliminate the control in the linear system and we obtain an algorithm which requires hardly any more work than for the unconstrained problem. The numerical results underscore this observation and show that the projected Newton method is an efficient solver for the phase field control problem with control constraints.

## 2. The Control Problem

Solid liquid phase transitions in pure materials can be modeled by so called phase field models. These models can be viewed as extensions of the classical Stefan problem. Instead of using an explicit condition for the interface boundary between



solid and liquid phase, phase field models consist of systems of two nonlinear parabolic equations. The interface can be constructed from the so called phase function. The advantage of phase field models is that the explicit use of the free boundary is avoided and that phenomena associated with surface tension and supercooling are incorporated into the model. In this paper we consider the phase field model proposed in [4], [7], [8]. Let  $\Omega$  be a subset of  $\mathbb{R}^n$ ,  $n \leq 3$ , with  $C^2$  boundary and let  $Q = \Omega \times (0, T)$ . By  $u(x, t)$  we denote the temperature at  $x \in \Omega$  and time  $t \in [0, T]$ . The phase function denoted by  $\varphi(x, t)$  describes the phase of the medium. The medium is purely solid if  $\varphi(x, t) = -1$  and it is purely liquid if  $\varphi(x, t) = +1$ . The boundary between the two phases can be described by

$$\Gamma(x, t) = \{(x, t) \in Q \mid \varphi(x, t) = 0\}.$$

The relation between the two quantities  $u$  and  $\varphi$  is given through a boundary value problem with two coupled semilinear parabolic differential equations,

$$(2.1) \quad \begin{aligned} u_t + \frac{\ell}{2}\varphi_t &= \kappa\Delta u + f \\ \tau\varphi_t &= \xi^2\Delta\varphi + g(\varphi) + 2u \end{aligned} \quad \text{on } Q = \Omega \times (0, T).$$

Here  $g$  is a nonlinear function of the type

$$g(z) = az + bz^2 - cz^3, \quad a, b, c \in C^1(\Omega), \quad a, b, c \geq 0.$$

Typically values like  $a = c = 0.5$ ,  $b = 0$ , are used in [4], [8]. The constants  $\kappa$ ,  $\ell$ ,  $\tau$ , and  $\xi$  denote the heat conductivity, the latent heat, the relaxation time, and the length scale of the interface, respectively. The function  $f$  represents the control input for the system. The boundary conditions are of Neumann type

$$(2.2) \quad \frac{\partial}{\partial n}u = 0, \quad \frac{\partial}{\partial n}\varphi = 0 \quad \text{on } \partial\Omega \times (0, T)$$

and the initial conditions are

$$(2.3) \quad u(x, 0) = u_0(x), \quad \varphi(x, 0) = \varphi_0(x) \quad x \in \Omega.$$

We assume that the initial conditions satisfy

$$(2.4) \quad u_0, \varphi_0 \in W_{\infty}^2(\Omega), \quad \frac{\partial}{\partial n}u_0 = \frac{\partial}{\partial n}\varphi_0 = 0.$$

Next we formulate the optimal control problem. Suppose that a certain desired phase function  $\varphi_d$  and a temperature  $u_d$  are given. The goal is to drive the melting process using  $f$  in such a way that the actual phase and temperature is as close as possible to the desired ones. We also include a cost term for the control  $f$  in the objective function  $J(f)$

$$(2.5) \quad J(f) = \frac{\alpha}{2}\|u - u_d\|_{L^2(Q)}^2 + \frac{\beta}{2}\|\varphi - \varphi_d\|_{L^2(Q)}^2 + \frac{\gamma}{2}\|f\|_{L^2(Q)}^2.$$

In this paper we also impose pointwise constraints on the control. Given  $f_{low}, f_{upp} \in L^\infty(Q)$  the set of admissible controls is defined by

$$F_{ad} = \{f \in L^\infty(Q) \mid f_{low} \leq f \leq f_{upp} \text{ on } Q\}.$$

This leads to the formulation of the optimal control problem using (2.1)–(2.5):

**Optimal Control Problem (OCP)**

*Find  $f^* \in F_{ad}$  such that  $J(f^*) \leq J(f)$  for all  $f \in F_{ad}$ .*

It has been shown in [11] that the state equation (2.1)–(2.3) has a unique solution and that the solution operator  $T(f) = (u, \varphi)$  corresponding to (2.1)–(2.3) satisfies

$$T : L^\infty(Q) \rightarrow W_p^{2,1}(Q) \times W_p^{2,1}(Q), \quad p < \infty.$$

The existence of optimal controls for the unconstrained problem has been shown in [11]. Analogous arguments can be used to prove existence of optimal controls for (OCP).

In the next section we look at the necessary optimality conditions which will lead to the construction of an efficient algorithm.

### 3. Necessary Optimality Conditions

If the objective function  $J(f)$  is Fréchet-differentiable then it is well known that the necessary optimality condition can be expressed as

$$(3.1) \quad J'(f^*)(f - f^*) \geq 0 \quad \text{for all } f \in F_{ad}.$$

It is shown in [11] that the solution map  $T$  is Fréchet-differentiable with Fréchet-derivative  $T'(f^*)(f) = (u, \varphi)$  given by the linearized system of (2.1):

$$(3.2) \quad \begin{aligned} u_t + \frac{1}{2}\varphi_t &= \kappa\Delta u + f \\ \tau\varphi_t &= \xi^2\Delta\varphi + g'(\varphi^*)\varphi + 2u \end{aligned} \quad \text{on } Q$$

with boundary conditions (2.2) and homogeneous initial conditions. In equation (3.2)  $(u^*, \varphi^*)$  denotes the solution of (2.1), (2.2), (2.3) with  $f = f^*$ , i.e.  $T(f^*) = (u^*, \varphi^*)$ . In order to write  $J'(f^*)f$  in a gradient notation we need to introduce the adjoint equation which is also a system of linear parabolic differential equations. For given  $f^*$  and  $T(f^*) = (u^*, \varphi^*)$  let  $(p^*, \psi^*)$  solve the system of adjoint equations:

$$(3.3) \quad \begin{aligned} -p_t^* &= \kappa\Delta p^* + 2\psi^* + \alpha(u^* - u_d) \\ -\tau\psi_t^* - \frac{1}{2}p_t^* &= \xi^2\Delta\psi^* + g'(\varphi^*)\psi^* + \beta(\varphi^* - \varphi_d) \end{aligned} \quad \text{on } Q,$$

with boundary and final conditions

$$(3.4) \quad \begin{aligned} \frac{\partial}{\partial n} p^* &= 0, \quad \frac{\partial}{\partial n} \psi^* = 0 \quad \text{on } \partial\Omega \times (0, T), \\ p^*(x, T) &= 0, \quad \psi^*(x, T) = 0 \quad \text{on } \Omega. \end{aligned}$$

Then we can show that the gradient  $\nabla J(f^*)$  can be computed by solving (2.1) and (3.3) and is given by

$$\nabla J(f^*) = p^* + \gamma f^*.$$

Note that due to the form of  $F_{ad}$  the necessary optimality condition (3.1) can be rewritten as

$$(3.5) \quad f^* = P(f^* - (p^* + \gamma f^*)),$$

where the projection  $P : L^\infty(Q) \rightarrow L^\infty(Q)$  is defined as

$$P(f)(x, t) = \begin{cases} f_{upp}(x, t) & \text{if } f(x, t) > f_{upp}(x, t), \\ f_{low}(x, t) & \text{if } f(x, t) < f_{low}(x, t), \\ f(x, t) & \text{else.} \end{cases}$$

If we use the variable transformation

$$u(x, t) \rightarrow u(x, t) + u_0(x), \quad \varphi(x, t) \rightarrow \varphi(x, t) + \varphi_0(x)$$

to obtain a problem with homogeneous initial and boundary conditions for  $u$  and  $\varphi$  and if we use (3.5), then we can write (3.1) as a system of nonlinear equations:

Find  $z = (u, \varphi, p, \psi, f)^T$  such that

**Necessary Optimality Conditions (NOC)**

$$\begin{aligned} u_t + \frac{l}{2}\varphi_t - \kappa\Delta u &= \kappa\Delta u_0 + f, \\ \tau\varphi_t - \xi^2\Delta\varphi &= \xi^2\Delta\varphi_0 + g(\varphi + \varphi_0) + 2(u + u_0) \\ -p_t - \kappa\Delta p &= 2\psi + \alpha(u + u_0 - u_d), \\ -\tau\psi_t - \frac{l}{2}p_t - \xi^2\Delta\psi &= g'(\varphi + \varphi_0)\psi + \beta(\varphi + \varphi_0 - \varphi_d), \\ f &= P(f - (p + \gamma f)) \\ \frac{\partial}{\partial n}u &= 0, \quad \frac{\partial}{\partial n}\varphi = 0, \quad \frac{\partial}{\partial n}p = 0, \quad \frac{\partial}{\partial n}\psi = 0 \quad \text{on } \partial\Omega \times (0, T), \\ u(x, 0) &= 0, \quad \varphi(x, 0) = 0, \quad p(x, T) = 0, \quad \psi(x, T) = 0 \quad \text{on } \Omega. \end{aligned}$$

To rewrite the problem (NOC) into a form more suitable for our approach, we introduce a linear operator  $S : L^p(Q)^4 \rightarrow W_p^{2,1}(Q)^4$ ,  $p > 1$ , which is the solution operator of a boundary value problem. For  $w = (w_1, w_2, w_3, w_4)^T \in L^p(Q)^4$ ,  $p > 1$ , let  $Sw = (u, \varphi, p, \psi)^T$  be the solution of

$$\begin{aligned} (3.6) \quad u_t + \frac{l}{2}\varphi_t - \kappa\Delta u &= w_1, \\ \tau\varphi_t - \xi^2\Delta\varphi &= w_2, \\ -p_t - \kappa\Delta p &= w_3, \\ -\tau\psi_t - \frac{l}{2}p_t - \xi^2\Delta\psi &= w_4, \end{aligned}$$

where the solution has to satisfy the homogeneous boundary, initial and final conditions from (NOC). Note that the system (3.6) is decoupled and  $Sw = (u, \varphi, p, \psi)^T$

can be computed by solving four simple parabolic equations. For example, the second equation (3.6) only depends on  $\varphi$ . If  $\varphi$  is computed, then it can be substituted into the first equation which then can be solved for  $u$ . Similarly, one can use the third and fourth equation to compute  $p$  and  $\psi$ . This block triangular structure of  $S$  will also be used for the numerical solution, cf. Section 5.

Later we need an extension  $S_c$  of  $S$  which is defined as

$$S_c(w, f) = (Sw, f), \quad (w, f) \in L^p(Q)^5.$$

If we define a nonlinear operator  $G : L^\infty(Q)^5 \rightarrow L^\infty(Q)^5$  which includes the nonlinearities of the differential equation and takes care of the inhomogeneous terms by

$$(3.7) \quad G(z) = \begin{pmatrix} \kappa \Delta u_0 + f \\ \xi^2 \Delta \varphi_0 + g(\varphi + \varphi_0) + 2(u + u_0) \\ 2\psi + \alpha(u + u_0 - u_d) \\ g'(\varphi + \varphi_0)\psi + \beta(\varphi + \varphi_0 - \varphi_d) \\ P(f - (p + \gamma f)) \end{pmatrix},$$

then the equation (NOC) can be rewritten as a fixed point equation to find  $z = (u, \varphi, p, \psi, f)^T$  such that

$$(3.8) \quad F(z) = z - S_c G(z) = 0$$

with a nonlinear operator  $F : X = L^\infty(Q)^5 \rightarrow L^\infty(Q)^5$ .

#### 4. A Projected Newton Method

For nonlinear problems of the type (3.8), Newton's method can be used for the numerical solution in a very efficient way, see e.g. [10]. A complicating factor is the appearance of the projection term in the last component of (3.8) besides the fact that systems of parabolic boundary value problems are included. The projection  $P$  destroys the Fréchet-differentiability of the operator  $F$ . Therefore we have to modify Newton's method in a proper way. A similar problem with ordinary differential equations has been considered in [12]. Here we extend this method to the case of partial differential equations.

If one considers only the last component of  $G$ , then a solution technique by successive approximation would lead to a method of the projected gradient. This is a first order method and converges rather slowly. The improvement of this method by using second order information is not straightforward. It was shown in [2] that a mere replacement of the gradient, in our case  $p + \gamma f$ , by a Newton or Newton-like direction does no longer guarantee descent and is therefore inappropriate. However, it was also shown in [2] that a fast convergent algorithm can be obtained if one considers for the computation of the Newton direction certain index sets which exclude subsets of the active indices. This was the motivation in [12] to restrict

the indices in the projection to the inactive one when using a Newton step for the solution of the nonlinear equation (3.8).

Following [12] we define the sets

$$\begin{aligned} A_{upp}(z) &= \{(x, t) \in Q \mid p(x, t) + \gamma f(x, t) > \|F(z)\|_X^r\}, \\ A_{low}(z) &= \{(x, t) \in Q \mid p(x, t) + \gamma f(x, t) < -\|F(z)\|_X^r\} \end{aligned}$$

depending on the size of the residual  $\|F(z)\|_X$  and an exponent  $r \in (0, 1)$ , and

$$A(z) = A_{upp}(z) \cup A_{low}(z), \quad I(z) = Q \setminus A(z).$$

At the optimal point  $z^* = (u^*, \varphi^*, p^*, \psi^*, f^*)^T$  it holds that  $F(z^*) = 0$  and, see (3.5),

$$\begin{aligned} p^*(x, t) + \gamma f^*(x, t) > 0 &\implies (x, t) \in A_{upp}^* = \{(x, t) \in Q \mid f^*(x, t) = f_{upp}(x, t)\}, \\ p^*(x, t) + \gamma f^*(x, t) < 0 &\implies (x, t) \in A_{low}^* = \{(x, t) \in Q \mid f^*(x, t) = f_{low}(x, t)\}. \end{aligned}$$

Thus,

$$A_{upp}(z^*) \subset A_{upp}^*, \quad A_{low}(z^*) \subset A_{low}^*.$$

This motivates to use  $A(z)$  as an estimate for the active set at the point  $z$ . However, this set is not necessarily equal to the true active set.

Since Newton's method is locally convergent we assume that the obvious choice to set  $f$  equal to the maximal and minimal value on  $A_{upp}(z)$ , and  $A_{low}(z)$ , respectively, is correct. Therefore, for given  $z = (u, \varphi, p, \psi, f)^T$  we reset

$$f = f_{upp} \text{ on } A_{upp}(z), \quad f = f_{low} \text{ on } A_{low}(z)$$

if these identities did not hold already. The main step is the computation of a correction of  $f$  on  $I(z)$  with the Newton step. However,  $I(z)$  is considered an approximation of the set of inactive indices. On this set the projection usually is the identity, so that we approximate on  $I(z)$  the nonlinear maps  $F$  and  $G$  by  $\tilde{F}$  and  $\tilde{G}$ , respectively.

$$\tilde{F}(z) = z - S_c \tilde{G}(z), \quad \tilde{G}(z) = \begin{pmatrix} \kappa \Delta u_0 + f \\ \xi^2 \Delta \varphi_0 + g(\varphi + \varphi_0) + 2(u + u_0) \\ 2\psi + \alpha(u + u_0 - u_d) \\ g'(\varphi + \varphi_0)\psi + \beta(\varphi + \varphi_0 - \varphi_d) \\ f - (p + \gamma f) \end{pmatrix}.$$

It can be shown [10] that  $\tilde{G}$  is Fréchet-differentiable and admits the representation

$$\tilde{G}'(z)(\delta z) = \begin{pmatrix} \delta f \\ g'(\varphi + \varphi_0)\delta\varphi + 2\delta u \\ 2\delta\psi + \alpha\delta u \\ g''(\varphi + \varphi_0)\delta\varphi + g'(\varphi + \varphi_0)\delta\psi + \beta\delta\varphi \\ \delta f - (\delta p + \gamma\delta f) \end{pmatrix}, \quad \delta z = \begin{pmatrix} \delta u \\ \delta\varphi \\ \delta p \\ \delta\psi \\ \delta f \end{pmatrix}.$$

Since  $S_c$  is linear we have that

$$\tilde{F}'(z)\delta z = \delta z - S_c \tilde{G}'(z)\delta z.$$

To describe the Newton step more precisely, let

$$Q_I z = (u, \varphi, p, \psi, \chi_I f)^T \text{ for } z = (u, \varphi, p, \psi, f)^T$$

with the characteristic function  $\chi_I$  on a set  $I \subset Q$ . Then the Newton step  $\delta z \in L^\infty(Q)^5$  is obtained from the solution of

$$(4.1) \quad Q_{I(z)} \tilde{F}'(z) Q_{I(z)} \delta z = -Q_{I(z)} \tilde{F}(z).$$

The equation (4.1) can be written as

$$Q_{I(z)} \delta z - Q_{I(z)} S_c \tilde{G}'(z) Q_{I(z)} \delta z = -Q_{I(z)} z + Q_{I(z)} S_c \tilde{G}(z)$$

or

$$(4.2) \quad Q_{I(z)}(z + \delta z) = Q_{I(z)} S_c [\tilde{G}'(z) Q_{I(z)}(z + \delta z) - \tilde{G}'(z) Q_{I(z)} z + \tilde{G}(z)].$$

With the definition of the solution operator  $S$  in (3.6) we can rewrite (4.2) as a system of boundary value problems. We denote the new iterate by a superscript  $+$ . The new iterate is given as the projection of  $z + Q_{I(z)} \delta z$  onto the feasible set:

$$z^+ = \tilde{P}(z + Q_{I(z)} \delta z),$$

where  $\tilde{P}(z) = (p, \psi, u, \varphi, P(f))^T$ . Since the projection is only applied to the fifth component, the first four components of the new iterate  $z^+$  are already obtained

from (4.2), which can be rewritten as

$$\begin{aligned}
 (4.3) \quad & u_t^+ + \frac{l}{2} \varphi_t^+ - \kappa \Delta u^+ = \kappa \Delta u_0 + \chi_{I(z)} \delta f + f, \\
 & \tau \varphi_t^+ - \xi^2 \Delta \varphi^+ = \xi^2 \Delta \varphi_0 + g'(\varphi + \varphi_0)(\varphi^+ - \varphi) \\
 & \quad + 2(u^+ + u_0) + g(\varphi + \varphi_0), \\
 & -p_t^+ - \kappa \Delta p^+ = 2\psi^+ + \alpha(u^+ + u_0 - u_d), \\
 & -\tau \psi_t^+ - \frac{l}{2} p_t^+ - \xi^2 \Delta \psi^+ = -g''(\varphi + \varphi_0)(\varphi^+ - \varphi)\psi^+ + g'(\varphi + \varphi_0)\psi^+ \\
 & \quad + \beta(\varphi^+ + \varphi_0 - \varphi_d), \\
 & \chi_{I(z)}(p^+ + \gamma(f + \delta f)) = 0,
 \end{aligned}$$

with homogeneous boundary and initial conditions of the type (2.2) and (3.4). The fifth component of  $z^+$  is given by  $f^+ = P(f + \chi_{I(z)} \delta f)$ , where  $\chi_{I(z)} \delta f$  is obtained as the fifth component of the solution of (4.3).

As in the unconstrained case we note that the last equation can be omitted by eliminating  $\chi_{I(z)} \delta f$ . From the last equation we obtain that

$$\chi_{I(z)} \delta f = -\chi_{I(z)} f - \frac{1}{\gamma} \chi_{I(z)} p^+,$$

which is substituted into the first equation in (4.3).

In summary, the conceptual algorithm looks as follows. Here we have included the inhomogeneous data again into the initial conditions.

### Algorithm

Given  $r \in (0, 1)$  and a current iterate  $z^r = (p^r, \varphi^r, u^r, \varphi^r, f^r)^T$ .

**Step 1** Determine the estimate for the active set

$$\begin{aligned}
 A_{\text{upp}}(z^r) &= \{(x, t) \in Q \mid p^r(x, t) + \gamma f^r(x, t) > \|F(z^r)\|_X^r\}, \\
 A_{\text{low}}(z^r) &= \{(x, t) \in Q \mid p^r(x, t) + \gamma f^r(x, t) < -\|F(z^r)\|_X^r\}, \\
 A^r &= A(z^r) = A_{\text{low}}(z^r) \cup A_{\text{upp}}(z^r), \\
 I^r &= I(z^r) = Q \setminus A(z^r),
 \end{aligned}$$

and overwrite  $f^r$  by

$$f^r(x, t) = \begin{cases} f_{\text{upp}}(x, t) & \text{on } A_{\text{upp}}(z^r), \\ f_{\text{low}}(x, t) & \text{on } A_{\text{low}}(z^r). \end{cases}$$

**Step 2** Solve the linear boundary value problem

$$\begin{aligned}
 u_t^+ + \frac{1}{2}\varphi_t^+ - \kappa\Delta u^+ &= \chi_{A^c} f^c - \frac{1}{\gamma}\chi_{I^c} p^+, \\
 \tau\varphi_t^+ - \xi^2\Delta\varphi^+ &= g'(\varphi^c)(\varphi^+ - \varphi^c) + 2u^+ + g(\varphi^c), \\
 -p_t^+ - \kappa\Delta p^+ &= 2\psi^+ + \alpha(u^+ - u_d), \\
 -\tau\psi_t^+ - \frac{1}{2}p_t^+ - \xi^2\Delta\psi^+ &= -g''(\varphi^c)(\varphi^+ - \varphi^c)\psi^c + g'(\varphi^c)\psi^+ + \beta(\varphi^+ - \varphi_d), \\
 u^+(x, 0) &= u_0(x), \quad \varphi^+(x, 0) = \varphi_0(x), \quad x \in \Omega, \\
 \frac{\partial}{\partial n} u^+(x, t) &= 0, \quad \frac{\partial}{\partial n} \varphi^+(x, t) = 0, \quad x \in \partial\Omega, \quad t \in (0, T), \\
 p^+(x, T) &= 0, \quad \psi^+(x, T) = 0, \quad x \in \Omega, \\
 \frac{\partial}{\partial n} p^+(x, t) &= 0, \quad \frac{\partial}{\partial n} \psi^+(x, t) = 0 \quad x \in \partial\Omega, \quad t \in (0, T).
 \end{aligned}$$

**Step 3** Set

$$f^+ = \chi_{A^c} f^c + P(\chi_{I^c}(f^c + \delta f^c)).$$

The system in Step 2 has been rewritten in a way that is very similar to the Newton system in the unconstrained case [10] and can be solved efficiently using multilevel approaches. A formulation of the system in Step 2 as a linear fixed point problem using the solution operator  $S$  yields a decoupling of the system and allows an efficient solution using multilevel approaches. The decoupling and the application of multilevel methods as well as further implementation details are described in more detail in the following section.

## 5. Implementation and Numerical Results

In each step of the projected Newton method introduced in the previous section we have to solve a system of four linear parabolic equations in the reduced set of variables  $\bar{z} = (u, \varphi, p, \psi)^T$ . With the solution operator  $S$  of (3.6) and

$$B = \begin{pmatrix} 0 & 0 & -\frac{1}{\gamma}\chi_{I^c} & 0 \\ 2 & g'(\varphi^c) & 0 & 0 \\ \alpha & 0 & 0 & 2 \\ 0 & -g''(\varphi^c)\psi^c + \beta & 0 & g'(\varphi^c) \end{pmatrix}, \quad b = \begin{pmatrix} \chi_{A^c} f^c \\ -g'(\varphi^c)\varphi^c + g(\varphi^c) \\ -\alpha u_d \\ +g''(\varphi^c)\varphi^c\psi^c - \beta\varphi_d \end{pmatrix},$$

the system arising in Step 2 of the algorithm can be written as

$$(5.1) \quad (I - K)\bar{z}^+ = b,$$

where  $K = SB$ . Due to the block triangular structure of the solution operator  $S$  the application of  $K$  to  $\bar{z}^+$  requires the sequential solution of a decoupled system of differential equations. For example, given  $\bar{z}^+ = (u^+, \varphi^+, p^+, \psi^+)^T$  the product  $(u, \varphi, p, \psi)^T = K\bar{z}^+$  can be evaluated by first solving the second equation in (3.6) with  $w_2 = 2u^+ + g'(\varphi^c)\varphi^+$  and corresponding boundary conditions. Then the solution  $\varphi$  can be inserted into the first equation and the first equation with



$w_1 = -\frac{1}{\gamma} \chi_{I^c} p^+$  can be solved for  $u$ . The third and fourth components  $p$  and  $\psi$  can be computed similarly. It should also be noted that for the sequential solution we only need to solve heat equations. There exist several efficient methods to perform this task numerically, such as multigrid methods, domain decomposition methods, or multilevel preconditioned conjugate gradient methods. If these methods are used, then the time needed to evaluate  $K\bar{z}^+$  for given  $\bar{z}^+$  is roughly linear in the number of variables introduced by the discretization.

Moreover, using the block triangular structure of  $S$ , the smoothing properties of the heat equation and Sobolev imbedding theorems one can show that  $K$  is a compact operator from  $L^\infty(Q)^4$  to  $L^\infty(Q)^4$ , [10]. Therefore (5.1) is a compact fixed point problem and one can use multigrid methods [1], [3], [9] for the solution of (5.1). In our implementations we use the multilevel method from [1], [3] which uses a fine and a coarse grid for the approximation of  $(I-K)^{-1}$ . Given an approximation  $\bar{z}^+$  for the solution of (5.1) the correction  $\delta\bar{z}^+$  yielding the exact solution  $\bar{z}^+ + \delta\bar{z}^+$  satisfies

$$(I - K)\delta\bar{z}^+ = b - (I - K)\bar{z}^+.$$

Using

$$(I - K)^{-1} = I + (I - K)^{-1}K \approx I + (I_l - K_l)^{-1}K,$$

where the index  $l$  denotes a coarse grid and  $(I_l - K_l)^{-1}$  is the inverse of a coarse grid operator, one finds that

$$\delta\bar{z}^+ \approx [I + (I_l - K_l)^{-1}K](b - (I - K)\bar{z}^+).$$

Thus the iteration used to solve the system of four linear parabolic equations in Step 2 of the algorithm is given by

$$(5.2) \quad \bar{z}^+ \leftarrow \bar{z}^+ + [I + (I_l - K_l)^{-1}K](b - (I - K)\bar{z}^+).$$

In our implementation we start this iteration using the current Newton iterate  $(u'', \varphi'', p'', \psi'')^T$ . One iteration requires the application of the fine grid operator  $K$ , which can be done very efficiently as explained previously and the solution of a coarse grid system with matrix  $(I_l - K_l)$ . Although the coarse grid system is considerably smaller than the fine grid system, it is still large. In the numerical experiment described below, the coarse grid system has 891 variables. Therefore it also has to be solved iteratively. Since  $K_l$  is nonsymmetric, we use GMRES [14]. In our numerical experiments one iteration (5.2) with starting value  $(u'', \varphi'', p'', \psi'')^T$  was sufficient to compute the solution  $\bar{z}^+ = (u^+, \varphi^+, p^+, \psi^+)^T$  with required accuracy.

For the discretization of the parabolic partial differential equations we use linear finite elements in space and backward Euler in time. The elliptic differential equations arising in each time step are solved using a hierarchical basis preconditioned conjugate gradient method [16].

The example we use to demonstrate the projected Newton multilevel method is taken from [6]. The following parameters and functions are used.

$$\Omega = (0, 1)^2, T = 0.05, a = \frac{1}{2}, b = 0, c = \frac{1}{2}, \kappa = \ell = 1, \tau = 0.05, \xi^2 = 10^{-2},$$

and

$$\begin{aligned} \varphi_d &= (x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 - \frac{3}{4} \left( (x - \frac{1}{2}) \cos \theta + (y - \frac{1}{2}) \sin \theta \right)^2 \\ &\quad - \left( R_0^2 + \frac{t}{T} (R_1^2 - R_0^2) \right), \\ u_d &= \frac{1}{2} (\tau \varphi_t - \xi_x^2 \varphi_{xx} - \xi_y^2 \varphi_{yy} + (\varphi^3 - \varphi)). \end{aligned}$$

where  $\theta = \frac{\pi}{4}(2 - \frac{t}{T})$ ,  $R_0 = 0.05$ ,  $R_1 = 0.15$ .

The weighting parameters in the objective function are chosen to be

$$\alpha = 50, \beta = 50, \gamma = 10^{-2}.$$

We made two tests runs. One for the unconstrained case  $f_{low} = -\infty$ ,  $f_{upp} = \infty$  and one for a constrained problem where we used  $f_{low} = -1.3$ ,  $f_{upp} = 1.3$ . In both cases we used a nested approach, starting on a coarse grid and using the interpolation of the coarse grid solution as a starting point on the next finer grid.

As mentioned previously, we use a finite element discretization in space and backward Euler in time. The triangulation of the spatial domain is obtained by dividing  $x$  and  $y$  axis into equidistant subintervals of length  $h$  and then dividing each of the resulting subsquares into two triangles. The grid is refined by doubling the number of subintervals in the  $x$  and  $y$  axis and doubling the number of time steps. Thus the number of triangles per time step increases by a factor of four and the total number of variables roughly increases by a factor of eight. We use hierarchical basis in our finite element spaces [16]. The level  $\ell = 0$  of the spatial grid consists of two triangles and four vertices. The process of increasing the level by one is described above and increases the number of triangles by a factor of four. For our discretization the level  $\ell$  spatial grid consists of  $2^{(2\ell+1)}$  triangles and  $(2^\ell + 1)^2$  vertices.

In the tables below  $\ell$  refers to the spatial grid level,  $ntime$  denotes the number of time steps (without the initial time), and  $node$  is the number of nodes in the space-time grid. The coarse grid is given by  $\ell = 3$ ,  $ntime = 10$ , which means that the number of nodes in the space-time grid is equal to  $(2^3 + 1)^2 \cdot 11 = 891$ . Since the variable in our problem is given by  $z = (u, \varphi, p, \psi, f)^T$ , the number of variables in the coarse grid is given by  $5 \cdot 891 = 4455$ .

The column  $itg$  shows the number of GMRES iterations needed to solve the coarse grid system in the multilevel approach. The times given show the accumulated time needed to solve the problem on the specified grid.

**Nested Multilevel Newton Method**  
(Unconstrained Problem,  $f_{low} = -\infty$ ,  $f_{upp} = \infty$ )

$\ell$	$n_{time}$	$n_{odes}$	$iter$	$\ F(z)\ $	$\ s\ $	$itg$	$time$
3	10	891	0	$0.608E-01$	$0.467E+00$	7	75
			1	$0.316E-01$	$0.351E+00$	9	
			2	$0.101E-01$	$0.593E-01$	8	
			3	$0.112E-02$	$0.108E-01$	12	
			4	$0.151E-03$			
4	20	6069	0	$0.145E+00$	$0.139E+00$	17	73
			1	$0.984E-02$			
5	40	44649	0	$0.448E-01$	$0.851E-01$	15	324
			1	$0.759E-02$			

TABLE 1

**Nested Multilevel Projected Newton Method**  
(Constrained Problem,  $f_{low} = -1.3$ ,  $f_{upp} = 1.3$ ,  $r = 0.9$ )

$\ell$	$n_{time}$	$n_{odes}$	$iter$	$\ F(z)\ $	$\ s\ $	$\ A(z)\ $	$itg$	$time$
3	10	891	0	$0.608E-01$	$0.467E+00$	10	7	77
			1	$0.316E-01$	$0.359E+00$	291	9	
			2	$0.102E-01$	$0.103E-01$	436	11	
			3	$0.152E-02$	$0.285E-01$	463	11	
			4	$0.110E-03$				
4	20	6069	0	$0.150E+00$	$0.140E+00$	18	17	72
			1	$0.987E-02$		2439		
5	40	44649	0	$0.475E-01$	$0.523E-01$	6990	17	354
			1	$0.624E-02$		18502		

TABLE 2

Figures 1 and 2 show the computed controls at time  $t = 0.05$ .

**Computed Control,  $f_{low} = -\infty, f_{upp} = \infty, t = 0.005$**

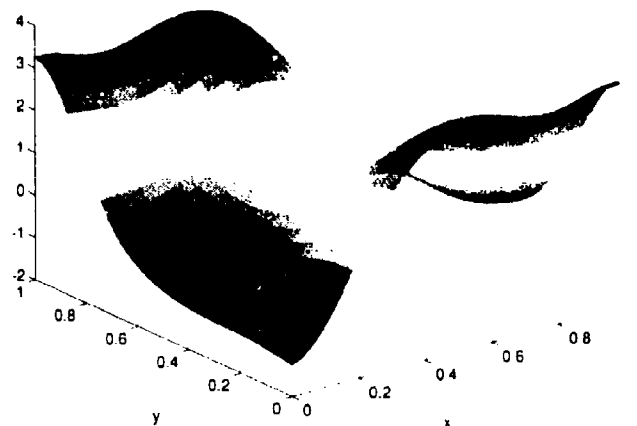


FIGURE 1

**Computed Control,  $f_{low} = -1.3, f_{upp} = 1.3, t = 0.005$**

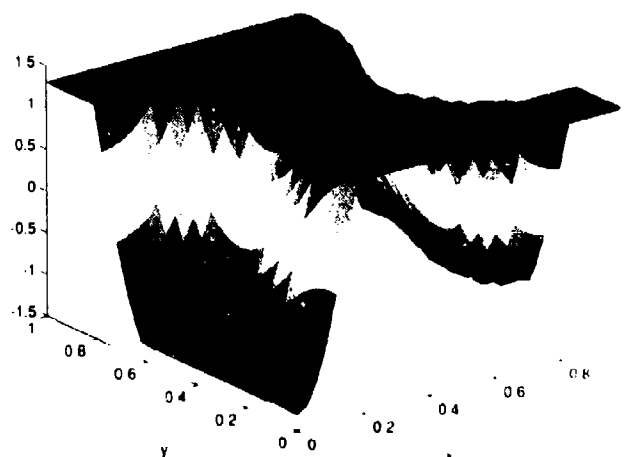


FIGURE 2

Figures 3 and 4 show the contours  $\{\psi = 0\}$  of the phase function for  $t = 0, 0.005, 0.01, \dots, 0.05$  computed in the unconstrained and constrained case. As noted in the introduction, these contours indicate the phase boundary.

### Contours of the Phase Function ,

$$(f_{\text{uncon}} - \infty, f_{\text{app}} - \infty)$$



FIGURE 4

### Contours of the Phase Function ,

$$(f_{\text{uncon}} - 1.11, f_{\text{app}} - 1.1)$$



FIGURE 3

**Acknowledgement.** The authors appreciate the help of Robert Scholz in the section on the numerical results.

## References

1. K. Atkinson, The numerical evaluation of fixed points for completely continuous operators, *SIAM J. Numer. Anal.*, **10**, 890-907, 1973.
2. D. H. Bertschke, Projected Newton methods for optimization problems with simple constraints, *SIAM J. Control Optim.*, **20**, 220-246, 1982.
3. H. Brakhage, Über die numerische Behandlung von Integralgleichungen nach der Quadraturformeltheorie, *Numer. Math.*, **2**, 183-196, 1969.
4. G. Caginalp, An analysis of a phase field model of a free boundary, *Arch. for Rational Mechanics and Analysis*, **92**, 205-245, 1986.
5. Z. Chen, *Das Phasefeldmodell für das Problem der Phasengrenzen*, *Analysis, Numerical and Scientific*, PhD thesis, Institut für Mathematik, Universität Augsburg, Germany, 1991.
6. Z. Chen and K. H. Hoffmann, Numerical solutions of the optimal control problem governed by a phase field model, In: W. Dorsch, F. Kappel, and K. Kunisch, editors, *Optimal Control of Partial Differential Equations*, Basel, Boston, Berlin, 1991, Birkhäuser Verlag.
7. J. W. Cahn and J. E. Levine, Diffusive interface model of diffusion-limited crystal growth, *Physica Rev. B*, **31**, 6119-6122, 1985.
8. G. Fox, Phase field models for free boundary problems, In: A. Fasano and M. Primicerio, editors, *Free Boundary Problems: Theory and Applications, Vol. II*, pages 580-589, Pitman, Boston, New York, Melbourne, 1983.
9. W. Hackbusch, *Multi-grid Methods and Applications*, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1985.
10. M. Henkelshaus, Numerical solution of a semilinear parabolic control problem, Technical report, Forschungsbericht, Universität Trier, FB IV - Mathematik, Postfach 3825, D-54286 Trier, 1993.
11. K. H. Hoffmann and L. Jiang, Optimal control problem of a phase field model for solidification, *Numer. Funct. Anal.*, **13**, 11-27, 1992.
12. C. T. Kelley and F. W. Sachs, Solution of optimal control problems by a pointwise projected Newton method, *Department of Mathematics, North Carolina State University, Raleigh, NC*, 1993.
13. G. Pinnau and P. C. Fife, Thermodynamically consistent models of phase field type for kinetics of phase transitions, *Physica D*, **43**, 41-62, 1990.
14. Y. Saad and M. H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Stat. Comp.*, **7**, 856-869, 1986.

15. A. A. Wheeler, B. F. Murray, and R. J. Schaefer, Computation of dendrites using a phase field model, *Physica D* 66:243-262, 1993.
16. H. Yserentant, On the multi-level splitting of finite element spaces, *Numer. Math.* 49:379-412, 1986.

M. Henkenschiess  
Interdisciplinary Center for Applied Mathematics  
and Department of Mathematics  
Virginia Polytechnic Institute and State University  
Blacksburg, VA 24061-0123, USA  
e-mail: ma.henkenschiess@cornell.edu

F. W. Sachs  
Lehrstuhl Mathematik  
Universität Trier  
D-54286 Trier, Federal Republic of Germany  
e-mail: sachs@uni-trier.de

## UNIFORM STABILIZABILITY OF NONLINEARLY COUPLED KIRCHHOFF PLATE EQUATIONS

MARY ANN HORN\* AND IRENA LASIECKA\*\*

School of Mathematics  
University of Minnesota

Department of Applied Mathematics  
University of Virginia

**ABSTRACT.** A system of two Kirchhoff plate equations with nonlinear coupling through both the boundary and the interior is considered. For this problem, it is proven that by appropriately choosing feedback controls, the energy of the system decays at a uniform rate. This result extends previous results in a number of directions: (i) it does not require any geometric hypotheses to be imposed on the domain; (ii) it allows for the presence of *nonlinear* coupling terms; (iii) it does not require the control functions to satisfy any growth conditions at the origin.

**1991 Mathematics Subject Classification.** 35J93

**Key words and phrases.** Uniform stabilization, boundary feedback, coupled plates, Kirchhoff plate

### 1. Introduction

As more results become available for single plates, attention has naturally shifted to more complex problems, such as those involving connected cores. Because one of the motivating factors in considering plate equations is to model more complicated vibrating flexible structures, our goal is to consider a coupled system, where the actions of one plate are allowed to affect those of the other.

The problem we consider is the following. Let  $\Omega$  be an open bounded domain in  $R^2$  with sufficiently smooth boundary,  $\Gamma$ . In  $\Omega$ , we consider the following system of nonlinearly coupled Kirchhoff's equations,

$$(1.1.a) \quad \begin{aligned} v_{tt} - \gamma^2 \Delta v_{tt} + \Delta^2 v &= k_1(u - v) & \text{in } Q_\infty = (0, \infty) \times \Omega \\ u_{tt} - \gamma^2 \Delta u_{tt} + \Delta^2 u &= k_2(v - u) \end{aligned}$$

\*This material is based upon work partially supported under a National Science Foundation Mathematical Sciences Postdoctoral Research Fellowship.

\*\*Partially supported by National Science Foundation Grant NSF DMS-9204338.



with initial conditions

$$(1.1.b) \quad \left. \begin{aligned} v(0, \cdot) &= v_0 \in H^2(\Omega), \quad v_t(0, \cdot) = v_1 \in H^1(\Omega) \\ u(0, \cdot) &= u_0 \in H^2(\Omega), \quad u_t(0, \cdot) = u_1 \in H^1(\Omega) \end{aligned} \right\} \quad \text{in } \Omega$$

and boundary conditions on  $\Sigma_\infty \equiv \Gamma \times (0, \infty)$ ,

$$(1.1.c) \quad \begin{aligned} \Delta v + (1 - \mu)\mathcal{B}_1 v &= g_1 \\ \frac{\partial}{\partial \nu} \Delta v + (1 - \mu)\mathcal{B}_2 v - \gamma^2 \frac{\partial}{\partial \nu} v_{tt} - v + \frac{\partial}{\partial \tau} h_1 \left( \frac{\partial}{\partial \tau} (v - u) \right) &= l_1(v - u) \\ &= g_2 - \frac{\partial}{\partial \tau} g_3 \end{aligned}$$

$$(1.1.d) \quad \begin{aligned} \Delta u + (1 - \mu)\mathcal{B}_1 u &= g_4 \\ \frac{\partial}{\partial \nu} \Delta u + (1 - \mu)\mathcal{B}_2 u - \gamma^2 \frac{\partial}{\partial \nu} u_{tt} - u + \frac{\partial}{\partial \tau} h_2 \left( \frac{\partial}{\partial \tau} (u - v) \right) &= l_2(u - v) \\ &= g_5 - \frac{\partial}{\partial \tau} g_6. \end{aligned}$$

Here,  $\gamma > 0$  is a constant proportional to the thickness of the plate,  $0 < \mu < \frac{1}{2}$  is Poisson's ratio, the operators  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are given by

$$(1.2) \quad \begin{aligned} \mathcal{B}_1 w &= 2n_1 n_2 w_{,\tau\tau} - n_1^2 w_{,\eta\eta} - n_2^2 w_{,\tau\tau} \\ \mathcal{B}_2 w &= \frac{\partial}{\partial \tau} [(n_1^2 - n_2^2) w_{,\tau\tau} + n_1 n_2 (w_{,\eta\eta} - w_{,\tau\tau})]. \end{aligned}$$

The functions  $k_i(s)$ ,  $h_i(s)$ ,  $l_i(s)$  are assumed to be differentiable with the property that for  $s \in \mathbb{R}$ ,

$$(1.1) \quad \begin{cases} s k_i(s) \rightarrow 0 \\ s h_i(s) \rightarrow 0 \\ s l_i(s) \rightarrow 0 \\ |h(s)| \leq C |s|^K \quad \text{for some } K > 0 \text{ and } |s| \rightarrow 1. \end{cases}$$

Control functions are denoted by  $g_i \in L_2(\Sigma)$  and physically represent moments and shears applied to the edge of the plate (see [5]). One could also consider an addition of nonlinear coupling in the first boundary condition in (1.1.c)-(1.1.d). Since this will not add extra difficulties, for simplicity we do not consider the additional coupling.

The control model for a single Kirchhoff plate (without coupling) was introduced and analyzed in [5]. Systems of equations of the type as in (1.1) are motivated by problems arising in the modelling of parallelly coupled plates (see [11], where a system of two one-dimensional wave equations with linear boundary conditions was considered). Indeed, parallel connection of plates introduces a force acting upon the system which is proportional to the difference of the displacement. If the plates are connected by, for example, springs, then linear theory gives the constants  $k$  and  $l$ . In the above system, a more general situation is assumed. Coupling of the plates is allowed to produce nonlinear forces. Also, we allow for the fact that the coupling may occur only on the boundary, in which case  $k_i = 0$  and  $l_i = 0$ .

The goal of this paper is to show that for appropriately chosen feedback controls, the energy of system (1.1) decays to zero at a uniform rate. Following

[5], we choose the nonlinear feedback controls of the form

$$(1.3) \quad \begin{aligned} g_1 &= -f_1\left(\frac{\partial}{\partial \nu} v_t\right) \\ g_2 &= g(v_t) \\ g_3 &= f_2\left(\frac{\partial}{\partial \tau} v_t\right) \\ g_4 &= -f_1\left(\frac{\partial}{\partial \nu} u_t\right) \\ g_5 &= g(u_t) \\ g_6 &= f_2\left(\frac{\partial}{\partial \tau} u_t\right). \end{aligned}$$

The following hypotheses are assumed on the nonlinear functions  $g$  and  $f_i$ : The functions  $g$  and  $f_i$  are continuous, monotone, zero at the origin and subject to the following growth conditions

$$(H-2) \quad \begin{cases} m|s| \leq |f_i(s)| \leq M|s| & \text{for } |s| > 1, \quad i = 1, 2 \\ |g(s)| \leq M|s|^p & \text{for } |s| > 1 \end{cases}$$

for some positive constants  $m$  and  $M$ .

To define the energy of the system, we introduce the bilinear form

$$(1.4) \quad \begin{aligned} a(w, v) &= \int_{\Omega} (\Delta w \Delta v + (1 - \mu)(2w_{xy}v_{xy} - w_{xx}v_{yy} - w_{yy}v_{xx})) d\Omega \\ &\quad + \int_{\Gamma} wv d\Gamma. \end{aligned}$$

The energy functional is defined by  $E(t) = E_v(t) + E_u(t)$ , where

$$(1.5) \quad E_v(t) = \frac{1}{2} \int_{\Omega} \{|w_t|^2 + \gamma^2 |\nabla w_t|^2\} d\Omega + \frac{1}{2} a(w, w).$$

The main result of our paper is the following.

**Theorem 1.1.** Assume (H-1) and (H-2) and consider equation (1.1) with (1.3).

I. For any initial data  $v_0, u_0 \in H^2(\Omega)$ ,  $v_1, u_1 \in H^1(\Omega)$ , there exists a solution  $(v, u) \in C(0, \infty; H^2(\Omega)) \cap C^1(0, \infty; H^1(\Omega))$  such that

$$(1.6) \quad \nabla v_t|_{\Gamma} \in L_2(\Sigma_{\infty}), \quad \nabla u_t|_{\Gamma} \in L_2(\Sigma_{\infty})$$

II. Assuming, in addition, that

$$(H-3) \quad \begin{cases} m \leq f'_i(s) \leq M \\ |g'(s)| \leq M(|s|^{p-1} + 1), \end{cases}$$

then there exists constants  $C, \omega > 0$  such that

$$(1.7) \quad E(t) \leq Ce^{-\omega t}, \quad t \geq 0,$$

where the constants  $C, \omega$  may depend on  $E(0)$ , but they are independent of  $t \geq 0$  and  $\gamma > 0$ .

Exponential decay rates for a system of one dimensional wave equations with distributed linear coupling and linear boundary dissipation was proven recently in [10] and [11]. The methods of [10] and [11] rely on spectral analysis which is typically restricted to one dimensional problems. Moreover, [10] and [11] state explicitly that other methods such as multipliers (see [3]) or direct methods (see [4]), which are tailored to multi-dimensional stability analysis, fail even in the case of linear coupling of a one dimensional wave equation with linear dissipation. We shall see that a different and more general approach based on intrinsic comparison of asymptotic behavior of the energy functional with the solution to a certain nonlinear ordinary differential equation problem (introduced in [7]) prove successful in this more complicated situation.

A critical role in the proof of Theorem 1.1 is played by a result on stabilizability of a semilinear Kirchhoff plate with nonlinear boundary conditions. Since this result is of independent interest, we shall state it below.

Let  $a_i$  (resp.,  $b$ ) be given elements in  $L_2(\Sigma_\infty)$  (resp.  $b \in C(0, \infty; L_{2r}(\Gamma))$ ). Consider the following equation of a nonautonomous semilinear Kirchhoff plate,

$$(1.8) \quad \begin{cases} w_{tt} - \gamma^2 \Delta w_{tt} + \Delta^2 w = -k(w) & \text{in } Q_\infty \\ w(0) = w_0 \in H^2(\Omega); \quad w_t(0) = w_1 \in H^1(\Omega) & \text{in } \Omega \\ \Delta w + (1-\mu)\mathcal{B}_1 w = [f_1(\frac{\partial}{\partial \nu} w_t + a_1) - f_1(a_1)] & \text{on } \Sigma_\infty \\ \frac{\partial}{\partial \nu} \Delta w + (1-\mu)\mathcal{B}_2 w = \gamma^2 \frac{\partial}{\partial \nu} w_{tt} - w + \frac{\partial}{\partial \nu} b(\frac{\partial}{\partial \nu} w) - l(w) \\ \quad g(w_t + b) - g(b) - \frac{\partial}{\partial \nu} [f_2(\frac{\partial}{\partial \nu} w_t + a_2) - f_2(a_2)] & \text{on } \Sigma_\infty. \end{cases}$$

Under the assumption that the functions  $k, n, l$  satisfy (H-1) and the functions  $g, f_i$  satisfy (H-2), one can show (by techniques identical to those in [7]), that for all initial data  $w_0 \in H^2(\Omega)$ ,  $w_1 \in H^1(\Omega)$ , there exists a solution  $w$  (not necessarily unique) to problem (1.8) such that

$$(1.9) \quad E_w(t) \in C^1(E_w(0)).$$

$$(1.10) \quad \nabla w|_{\Sigma_\infty} \in L_2(\Sigma_\infty).$$

**Theorem 1.2.** Assume that (H-1) and (H-2) hold true. Then there exist constants  $C, \omega > 0$  such that every solution to (1.8) of finite energy satisfies

$$(1.11) \quad E_w(t) \leq C e^{-\omega t}.$$

The constants  $C$  and  $\omega$  may depend on the initial energy  $E_w(0)$  and  $\|b\|_{C(0, \infty; L_{2r}(\Gamma))}$ .

In the special case when  $a_i = b = 0$  (or they are constant), a more general version of Theorem 1.2 holds true.

**Theorem 1.3.** Let  $a_i = b = 0$ . Assume (H-1) and (H-2). Then every solution of finite energy which satisfies (1.10) obeys the following estimate. For some  $T_0 > 0$ ,

$$(1.12) \quad E_w(t) \leq S(\frac{t}{T_0} + 1) \text{ for } t \geq T_0.$$

where  $\mathcal{S}(t) \rightarrow 0$  as  $t \rightarrow \infty$  and is the solution (contraction semigroup) of the differential equation

$$(1.13) \quad \begin{cases} \frac{d}{dt}\mathcal{S}(t) + q(\mathcal{S}(t)) = 0 \\ \mathcal{S}(0) = E_w(0), \end{cases}$$

and  $q(x)$  is a strictly monotone function constructed in terms of  $g$  and  $f_i$  (see Remark below).

*Remark 1.1:* To construct the function  $q(x)$ , we proceed as follows: Let the function  $h(x)$  be defined by:

$$(1.14) \quad h(x) \equiv h_1(x) + h_2(x),$$

where  $h_i(x)$  are concave, strictly increasing functions with  $h_i(0) = 0$  such that

$$(1.15) \quad h_i(s f_i(s)) \geq s^2 + f_i^2(s) \quad |s| \leq 1 \quad i = 1, 2.$$

(Such functions can be easily constructed. See [7].) Then  $h(x)$  enjoys the same properties, i.e., it is concave, strictly increasing, and  $h(0) = 0$ . Define

$$(1.16) \quad \tilde{h}(x) \equiv h\left(\frac{x}{mcs \Sigma_F}\right).$$

Since  $\tilde{h}$  is monotone increasing, for every  $c \geq 0$ ,  $cI + \tilde{h}$  is invertible. Setting

$$(1.17) \quad p(x) \equiv (cI + \tilde{h})^{-1}(Kx),$$

where  $c = \frac{1}{mcs \Sigma_F}(m^{-1} + M)$ , the constant  $K$  will generally depend on  $E_w(0)$  unless  $r = 1$  and  $h, k, l$  are linearly bounded. We then define  $q(x)$  by

$$(1.18) \quad q(x) \equiv x - (I + p)^{-1}(x), \quad x > 0.$$

*Remark 1.2:* In the special case when the growth at the origin of the nonlinear boundary feedbacks is specified, one can compute explicitly the governing decay rates for the energy function,  $E_w(t)$ . Indeed, this can be easily accomplished by constructing the appropriate function  $q$  and solving the nonlinear ODE problem, (1.13). For instance, if the nonlinear functions  $f_i$  and  $g$  have a linear growth at the origin, then (1.12) specializes to

$$(1.19) \quad E_w(t) \leq C(E_w(0))e^{-\omega t} \quad \text{for some } \omega > 0.$$

If, instead, the nonlinearities are of polynomial growth at the origin (e.g.  $\sim x^p$ ,  $p > 1$ ), then  $E_w(t) \leq C(E_w(0))t^{-\frac{2}{p-1}}$ .

*Remark 1.3:* The result of Theorem 1.3 should be compared to a recent result on uniform decay rates obtained in [12] (and references therein) for a linear Euler-Bernoulli model with nonlinear boundary dissipation. Indeed, in [12], the uniform (polynomial) decay rates were obtained under the following hypotheses: (i)  $\Omega$  is star-shaped, (ii) nonlinear functions  $f_i$  and  $g$  are subject to a linear growth at infinity and polynomial growth at the origin (hypotheses (1.25), (1.26) in [12]),

(iii)  $k = 0$ ,  $h = 0$ , (iv)  $l$  is linear. Under the above hypotheses, it is shown in [12] that *smooth*<sup>1</sup> solutions decay polynomially to zero.

In view of this, Theorem 1.3 extends the above result in several directions: (i) it does not require any geometric hypotheses to be imposed on  $\Omega$ , (ii) the result of Theorem 1.3 remains valid for all weak solutions, (iii) it allows the presence of nonlinear terms  $h, l$  on the boundary together with the nonlinear interior term  $k$ , and, finally, (iv) it does not require any growth conditions at the origin.

Needless to say, the techniques leading to the proof of Theorem 1.3 are very different (unfortunately more complicated) from those in [12]. Indeed, Lyapunov's function method used in [12] runs into well recognized difficulties when dealing with the level of generality presented in Theorem 1.3.

The outline of our paper is as follows. In section 2, we prove the result of Theorem 1.2 which is critical to the proof of Theorem 1.1 in section 4. The proof of Theorem 1.3 is relegated to section 3.

## 2. Proof of Theorem 1.2

### 2.1 Preliminaries

Our goal is to prove energy decay rates for problem (1.8). In order to do this, one needs to perform certain partial differential equation calculations on the problem. These calculations require regularity of the solutions higher than is available. Since our nonlinear problem may not have a sufficiently regular solution (even if the initial data are smooth), we resort to an approximation argument (this argument was used in the context of wave equations in [7]). In fact, the idea here is to approximate solutions to the nonlinear problem (1.8) by solutions to different (linear) problems. Since this linear problem admits regular solutions for smooth initial data, the partial differential equation calculations can be performed on this problem. Final passage to the limit on the approximation problem allows us to obtain needed energy identities for the original nonlinear problem.

To follow our program, we start by defining appropriate approximations for the quantities in equation (1.8). To do this, we introduce the following notation:

$$(2.1) \quad \hat{f}_i(s) \equiv f_i(s + a_i) - f_i(a_i); \quad \hat{g}(s) \equiv g(s + b) - g(b).$$

Note that by virtue of our hypothesis (H-3),  $\hat{f}_i$  and  $\hat{g}$  are strictly monotone functions and

$$(2.2) \quad m \leq \hat{f}'_i(s) \leq M.$$

<sup>1</sup>A smoothness assumption was not stated in [12], but implicitly assumed (to validate arguments leading to decay estimates).

**Corollary 2.1.** *Let  $w$  be a solution to (1.8) such that (1.9) and (1.10) hold true. Let  $T > 0$ . Then*

$$(2.3) \quad k(w) \in L_2(0, T; L_2(\Omega)),$$

$$(2.4) \quad \frac{\partial}{\partial \tau} h \left( \frac{\partial}{\partial \tau} w \right) \in L_2(0, T; H^{-1}(\Gamma)); \quad l \in L_2(0, T; L_2(\Gamma)),$$

$$(2.5) \quad \hat{f}_1 \left( \frac{\partial}{\partial \nu} w_t \right) \in L_2(0, T; L_2(\Gamma)),$$

$$(2.6) \quad \hat{g}(w_t) - \frac{\partial}{\partial \tau} \hat{f}_2 \left( \frac{\partial}{\partial \tau} w_t \right) \in L_2(0, T; H^{-1}(\Gamma)).$$

**Proof of Corollary 2.1:** Regularity in (2.3), (2.4) follows using Sobolev's Imbeddings from (1.9), (H-1).

Hypothesis (H-3) together with (1.10) implies

$$(2.7) \quad \begin{aligned} f_1 \left( \frac{\partial}{\partial \nu} w_t \right) &\in L_2(\Sigma_T) \\ f_2 \left( \frac{\partial}{\partial \tau} w_t \right) &\in L_2(\Sigma_T). \end{aligned}$$

Hence,

$$(2.8) \quad \frac{\partial}{\partial \tau} f_2 \left( \frac{\partial}{\partial \tau} w_t \right) \in L_2(0, T; H^{-1}(\Gamma)).$$

On the other hand, with  $\phi \in L_2(0, T; H^1(\Gamma))$ ,

$$(2.9) \quad \begin{aligned} \int_0^T \int_{\Gamma} |g(w_t(t, x)) \phi(t, x)| dx dt &\leq C \int_0^T \|\phi(t)\|_{H^1(\Gamma)} \int_{\Gamma} |w_t(t, x)|^r dx dt \\ &\leq C \int_0^T \|\phi(t)\|_{H^1(\Gamma)} \|w_t(t)\|_{L_p(\Gamma)}^r dt \\ &\leq C \int_0^T \|\phi(t)\|_{H^1(\Gamma)} E_w(t)^{r/2} dt \\ &\leq C E_w(0)^{r/2} \int_0^T \|\phi(t)\|_{H^1(\Gamma)} dt, \end{aligned}$$

where the first inequality follows from hypothesis (H-2), the second and third follow from Sobolev Imbeddings and the boundedness of  $\Gamma$ , and the fourth from trace theory. Hence,

$$(2.10) \quad g(w_t) \in L_2(0, T; H^{-1}(\Gamma)),$$

which, together with (2.8) proves (2.6).  $\square$

Let  $w$  be the solution of the original problem (1.8). By using the regularity properties in (1.10), (2.3)-(2.6), along with density of approximate (see below)

Sobolev spaces, we are in a position to define

$$(2.11) \quad f_n \in H^{1,1}(Q_T); \quad \|f_n + k(w)\|_{L_2(0,T;L_2(\Omega))} \longrightarrow 0$$

$$(2.12) \quad f_{1n} \in H^{1,1}(\Sigma_T); \quad \|f_{1n} - \hat{f}_1(\frac{\partial}{\partial \nu} w_t)\|_{L_2(\Sigma_T)} \longrightarrow 0$$

$$(2.13) \quad f_{2n} \in H^{1,1}(\Sigma_T); \quad \|f_{2n} - [\dot{g}(w_t) - \frac{\partial}{\partial \tau} \hat{f}_2(\frac{\partial}{\partial \tau} w_t) - \frac{\partial}{\partial \tau} h(\frac{\partial}{\partial \tau} w) + l(w)]\|_{L_2(0,T;H^{-1}(\Gamma))} \longrightarrow 0$$

$$(2.14) \quad \alpha_n \in H^{1,1}(\Sigma_T); \quad \|\alpha_n - \frac{\partial}{\partial \nu} w_t\|_{L_2(\Sigma_T)} \longrightarrow 0$$

$$(2.15) \quad \beta_n \in H^{1,1}(\Sigma_T); \quad \|\beta_n - (w_t - \frac{\partial^2}{\partial \tau^2} w_t)\|_{L_2(0,T;H^{-1}(\Gamma))} \longrightarrow 0,$$

where  $Q_T \equiv \Omega \times (0, T)$  and  $\Sigma_T \equiv \Gamma \times (0, T)$ . We consider the following approximating problem:

$$(2.16) \quad \begin{cases} w_{n,t} - \gamma^2 \Delta w_{n,t} + \Delta^2 w_n = f_n \\ w_n(0) = w_{n,0}; \quad w_{n,t}(0) = w_{n,1} \\ \Delta w_n + (1-\mu)\mathcal{B}_1 w_n + \frac{\partial}{\partial \nu} w_{n,t}|_{\Gamma} = -f_{1n} + \alpha_n \\ \frac{\partial}{\partial \nu} \Delta w_n + (1-\mu)\mathcal{B}_2 w_n - \gamma^2 \frac{\partial}{\partial \nu} w_{n,t} - w_n - w_{n,t} + \frac{\partial^2}{\partial \tau^2} w_{n,t}|_{\Gamma} \\ \quad = f_{2n} - \beta_n. \end{cases}$$

where

$$(2.17) \quad \|w_{n,0} - w_0\|_{H^2(\Omega)} \rightarrow 0; \quad \|w_{n,1} - w_1\|_{H^1(\Omega)} \rightarrow 0,$$

and  $(w_{n,0}, w_{n,1}) \in \mathcal{D}$ , where  $\mathcal{D}$ , as dense set of  $\mathcal{H}$ , consists of  $w_{n,0} \in H^4(\Omega)$ ,  $w_{n,1} \in H^3(\Omega)$ , where  $w_{n,0}, w_{n,1}$  satisfy the appropriate compatibility conditions on the boundary. By standard linear semigroup methods, one easily shows that the linear problem, (2.16), admits a classical solution.

$$(2.18) \quad w_n \in C(0, T; H^4(\Omega)) \cap C^1(0, T; H^3(\Omega)).$$

The following proposition plays a critical role in our development.

**Proposition 2.1.** *Let  $w_n$  (respectively,  $w$ ) be a solution of (2.16) (respectively, (1.8)). Then as  $n \rightarrow \infty$ , the following convergence holds.*

$$(2.19) \quad w_n \rightarrow w \text{ in } C(0, T; H^2(\Omega)) \cap C^1(0, T; H^1(\Omega))$$

$$(2.20) \quad \nabla w_{n,t}|_{\Gamma} \rightarrow \nabla w_t \text{ in } L_2(\Sigma_T).$$

**Proof:** Consider the equation satisfied by the difference  $w_n - w_m$ . Taking the inner product of this equation with  $w_{n,t} - w_{m,t}$  and integrating the result from 0 to  $T$  yields

$$(2.21) \quad \begin{aligned} & E_{w_n - w_m}(T) + \int_0^T \int_{\Gamma} [\dot{w}_t^2 + |\nabla \dot{w}_t|^2] d\Gamma dt \\ &= \int_0^T \int_{\Omega} (f_n - f_m) \dot{w}_t d\Omega dt + \int_0^T \int_{\Gamma} (f_{1n} + \alpha_n - f_{1m} - \alpha_m) \frac{\partial}{\partial \nu} \dot{w}_t d\Gamma dt \\ &\quad + \int_0^T \int_{\Gamma} (f_{2n} + \beta_n - f_{2m} - \beta_m) \dot{w}_t d\Gamma dt + E_{w_n - w_m}(0), \end{aligned}$$

where  $\dot{w} \equiv w_n - w_m$ . Hence,

$$(2.22) \quad \begin{aligned} & C_0 \|\dot{w}(T)\|_{H^2(\Omega)}^2 + \|\dot{w}_t(T)\|_{H^1(\Omega)}^2 + \|\nabla \dot{w}_t\|_{L_2(\Sigma_T)}^2 + \|\dot{w}_t\|_{L_2(\Sigma_T)}^2 \\ &\leq \frac{1}{2} \|f_n - f_m\|_{L_2(0, T; H^{-1}(\Omega))}^2 + \frac{1}{2} \int_0^T \|\dot{w}_t(t)\|_{H^1(\Omega)}^2 dt \\ &\quad + \frac{1}{2} \|f_{1n} - \alpha_n - f_{1m} + \alpha_m\|_{L_2(\Sigma_T)}^2 + \frac{1}{2} \left\| \frac{\partial}{\partial \nu} \dot{w}_t \right\|_{L_2(\Sigma_T)}^2 \\ &\quad + \frac{1}{2} \|f_{2n} - \beta_n - f_{2m} + \beta_m\|_{L_2(0, T; H^{-1}(\Omega))}^2 \\ &\quad + \frac{1}{2} \|\dot{w}_t\|_{L_2(0, T; H^1(\Omega))}^2 + E_{w_n - w_m, 1}(0), \end{aligned}$$

and

$$(2.23) \quad \begin{aligned} & \|\dot{w}(T)\|_{H^2(\Omega)}^2 + \|\dot{w}_t(T)\|_{H^1(\Omega)}^2 + \|\nabla \dot{w}_t\|_{L_2(\Sigma_T)}^2 \\ &\leq C [\|f_n - f_m\|_{L_2(0, T; H^{-1}(\Omega))}^2 + \|f_{1n} - f_{1m}\|_{L_2(\Sigma_T)}^2 \\ &\quad + \|\alpha_n - \alpha_m\|_{L_2(\Sigma_T)}^2 + \|f_{2n} - f_{2m}\|_{L_2(0, T; H^{-1}(\Omega))}^2 \\ &\quad + \|\beta_n - \beta_m\|_{L_2(0, T; H^{-1}(\Omega))}^2 + E_{w_n - w_m, 1}(0)] \\ &\longrightarrow 0 \text{ as } n \rightarrow \infty, \end{aligned}$$

where the limit follows by using (2.11)-(2.15). Thus, by (2.8) and Corollary 2.1,

$$(2.24) \quad \begin{aligned} & w_n \rightarrow w^* \text{ in } C(0, T; H^2(\Omega)) \cap C^1(0, T; H^1(\Omega)) \\ & \nabla w_{n,t}|_{\Gamma} \rightarrow \nabla w_t^*|_{\Gamma} \text{ in } L_2(\Sigma_T). \end{aligned}$$

This allows us to pass with the limit on the linear equation, (2.16). We obtain

$$(2.25) \quad \begin{cases} w_{tt}^* - \gamma^2 \Delta w_{tt}^* + \Delta^2 w^* + k(w) = 0 \\ w^*(0) = w_0; \quad w_t^*(0) = w_1 \in H^2(\Omega) \\ \Delta w^* + (1 - \mu) B_1 w^* + \frac{\partial}{\partial \nu} w_t^*|_{\Gamma} = -\hat{f}_1 \left( \frac{\partial}{\partial \nu} w_t \right) + \frac{\partial}{\partial \nu} w_t \\ \frac{\partial}{\partial \nu} \Delta w^* + (1 - \mu) B_2 w^* - \gamma^2 \frac{\partial}{\partial \nu} w_{tt}^* - w^* - w_t^* + \frac{\partial^2}{\partial \tau^2} w_t^*|_{\Gamma} \\ \quad = \hat{g}(w_t) - \frac{\partial}{\partial \tau} \hat{f}_2 \left( \frac{\partial}{\partial \tau} w_t \right) - w_t + \frac{\partial^2}{\partial \tau^2} w_t + l(w) - \frac{\partial}{\partial \tau} h \left( \frac{\partial}{\partial \tau} w \right). \end{cases}$$

Since  $w$  satisfies (2.25) and the solution to (2.25) is unique, we infer that  $w \equiv w^*$  and

$$(2.26) \quad \begin{aligned} & w_n \rightarrow w \text{ in } C(0, T; H^2(\Omega)) \cap C^1(0, T; H^1(\Omega)) \\ & \nabla w_{n,t}|_{\Gamma} \rightarrow \nabla w_t|_{\Gamma} \text{ in } L_2(\Sigma_T), \end{aligned}$$

as desired.  $\square$

Now we are in a position to prove the fundamental energy relation for problem (1.8).



**Lemma 2.1 (Energy Identity).** *Let  $w$  be the solution to (1.8). Then the following energy identity holds*

$$(2.27) \quad \hat{E}_w(t) - \hat{E}_w(0) + \int_{\Sigma_T} [\hat{g}(w_t)w_t + \hat{f}_1(\frac{\partial}{\partial\nu}w_t)\frac{\partial}{\partial\nu}w_t + \hat{f}_2(\frac{\partial}{\partial\tau}w_t)\frac{\partial}{\partial\tau}w_t]d\Gamma dt = 0,$$

where

$$\hat{E}_w(t) \equiv E_w(t) + \int_{\Omega} \mathcal{K}(w(t))d\Omega + \int_{\Gamma} [\mathcal{H}(\frac{\partial}{\partial\tau}w(t)) + \mathcal{L}(w(t))]d\Gamma,$$

and  $\mathcal{K}$ ,  $\mathcal{H}$  and  $\mathcal{L}$  are antiderivatives of  $k$ ,  $h$  and  $l$ .

**Proof:** We first prove this energy identity for the solution,  $w_n$ , of the approximation problem, (2.16). Indeed, by applying a standard energy argument (valid due to the smoothness of solutions  $w_n$ ) to (2.16), we obtain

$$(2.28) \quad \begin{aligned} E_{w_n}(T) - E_{w_n}(0) &+ \int_{\Sigma_T} |\frac{\partial}{\partial\nu}w_{n,t}|^2 d\Gamma dt \\ &+ \int_{\Sigma_T} (w_{n,t})^2 d\Gamma dt + \int_{\Sigma_T} |\frac{\partial}{\partial\tau}w_{n,t}|^2 d\Gamma dt \\ &= \int_0^T \int_{\Omega} f_n w_{n,t} d\Omega dt - \int_{\Sigma_T} (f_{1n} - \alpha_n) \frac{\partial}{\partial\nu}w_{n,t} d\Gamma dt \\ &\quad - \int_{\Sigma_T} (f_{2n} - \beta_n) w_{n,t} d\Gamma dt. \end{aligned}$$

Using convergence properties (2.11)-(2.15) and the result of Proposition 2.1, we obtain

$$(2.29) \quad \begin{aligned} E_w(T) - E_w(0) &+ \int_{\Sigma_T} |\frac{\partial}{\partial\nu}w_t|^2 d\Gamma dt \\ &+ \int_{\Sigma_T} (w_t)^2 d\Gamma dt + \int_{\Sigma_T} |\frac{\partial}{\partial\tau}w_t|^2 d\Gamma dt \\ &= - \int_0^T \int_{\Omega} k(w)w_t d\Omega dt + \int_{\Sigma_T} [-\hat{f}_1(\frac{\partial}{\partial\nu}w_t) + \frac{\partial}{\partial\nu}w_t] \frac{\partial}{\partial\nu}w_t d\Gamma dt \\ &\quad - \int_{\Sigma_T} [\hat{g}(w_t) - \frac{\partial}{\partial\tau}\hat{f}_2(\frac{\partial}{\partial\tau}w_t)]w_t d\Gamma dt + \int_{\Sigma_T} (w_t^2 + |\frac{\partial}{\partial\tau}w_t|^2) d\Gamma dt \\ &\quad - \int_{\Sigma_T} [l(w)w_t + h(\frac{\partial}{\partial\tau}w)\frac{\partial}{\partial\tau}w_t] d\Gamma dt. \end{aligned}$$

After canceling boundary terms and taking into account the definition of  $\hat{E}_w(t)$ , we obtain (2.27).

## 2.2 A Priori Estimates

To prove Theorem 1.2, we first show the following inequality holds.

**Theorem 2.1.** *Let  $w$  be the solution to (1.8) and  $T$  be sufficiently large. Then there exist constants,  $C > 0$  and  $C_T(E_w(0)) > 0$ , such that the following inequality holds:*

$$(2.30) \quad \begin{aligned} \int_0^T \hat{E}_w(t)dt - C\hat{E}_w(T) \\ \leq C_T(E_w(0), \|b\|_{C([0,T]; L_{2^*(\Gamma))}}) \int_{\Sigma_T} (\hat{g}(w_t)w_t + \hat{f}_1(\frac{\partial}{\partial\nu}w_t)\frac{\partial}{\partial\nu}w_t \\ + \hat{f}_2(\frac{\partial}{\partial\tau}w_t)\frac{\partial}{\partial\tau}w_t) d\Gamma dt. \end{aligned}$$

### 2.2.1 Estimates for the Approximated Problem

To prove Theorem 2.1, we begin by using a multiplier method on the approximation problem (2.16) to prove the following preliminary estimate.

**Lemma 2.2.** *Let  $(w_0, w_1) \in \mathcal{D}$  and  $0 < \alpha < T/2$ . Then the energy of system (2.16) satisfies the following estimate:*

$$(2.31) \quad \begin{aligned} & \int_0^{T-\alpha} E_{w_n}(t) dt \leq C_1(1 + \gamma^2) E_{w_n,1}(T - \alpha) + C_2(1 + \gamma^2) E_{w_n,1}(\alpha) \\ & \leq C_{T,\alpha,\gamma} \{ \|f_{1n}\|_{L_2(\Sigma_T)}^2 + \|f_{2n}\|_{L_2(\partial\Omega \times [0, T-\alpha])}^2 + \|\alpha_n - \frac{\partial}{\partial\nu} w_{n,t}\|_{L_2(\Sigma_T)}^2 \\ & \quad + \|\beta_n - w_{n,t} + \frac{\partial^2}{\partial\tau^2} w_{n,t}\|_{L_2(\partial\Omega \times [0, T-\alpha])}^2 + \|w_{n,t}\|_{L_2(\Sigma_T)}^2 \\ & \quad + (1 + \gamma^2) \|\nabla w_{n,t}\|_{L_2(\Sigma_T)}^2 + \|f_n\|_{L_2(\partial\Omega \times [0, T-\alpha])}^2 + l.o.(w_n) \}, \end{aligned}$$

where

$$(2.32) \quad l.o.(w_n) \equiv \|w_n\|_{L_2(\partial\Omega \times [0, T-\alpha])}^2,$$

$0 < \epsilon < 1/2$ , and  $\tilde{h} \equiv x - x_0$  for some  $x_0 \in \mathbb{R}^2$ .

**Proof of Lemma 2.2: Step 1: Identities.** From [5] (p. 84, (4.5.17)), with adjustments to take both the nonhomogeneous right-hand side of (2.16) and the boundary conditions into account, we have

$$(2.33) \quad \begin{aligned} & \int_0^T E_{w_n,1}(t) dt \\ & \leq -\frac{1}{2} [(w_{n,t}, \tilde{h} \cdot \nabla w_n)_{L_2(\Omega)} + \gamma^2 (\nabla w_{n,t}, \nabla(\tilde{h} \cdot \nabla w_n))_{L_2(\Omega)}]_0^T \\ & \quad + \frac{1}{2} [(w_{n,t}, w_n)_{L_2(\Omega)} + \gamma^2 (\nabla w_{n,t}, \nabla w_n)_{L_2(\Omega)}]_0^T \\ & \quad + \frac{1}{2} (\frac{\partial}{\partial\nu} w_n, f_{1n})_{L_2(\Sigma_T)} + \frac{1}{2} (w_n, f_{2n})_{L_2(\Sigma_T)} \\ & \quad + \frac{1}{2} (\frac{\partial}{\partial\nu} w_n, \alpha_n - \frac{\partial}{\partial\nu} w_{n,t})_{L_2(\Sigma_T)} + \frac{1}{2} (w_n, \beta_n - w_{n,t} + \frac{\partial^2}{\partial\tau^2} w_{n,t})_{L_2(\Sigma_T)} \\ & \quad - \int_{\Sigma_T} [\frac{\partial}{\partial\nu} (\tilde{h} \cdot \nabla w_n) f_{1n} + (\tilde{h} \cdot \nabla w_n) f_{2n}] d\Gamma dt \\ & \quad + \int_{\Sigma_T} [\frac{\partial}{\partial\nu} (\tilde{h} \cdot \nabla w_n) (\alpha_n - \frac{\partial}{\partial\nu} w_{n,t}) \\ & \quad - (\tilde{h} \cdot \nabla w_n) (\beta_n - w_{n,t} + \frac{\partial^2}{\partial\tau^2} w_{n,t})] d\Gamma dt \\ & \quad - \int_{\Sigma_T} (\tilde{h} \cdot \nabla w_n) w_n d\Gamma dt + \frac{1}{2} \int_{\Sigma_T} \tilde{h} \cdot \nu (w_{n,t}^2 + \gamma^2 |\nabla w_{n,t}|^2) d\Gamma dt \\ & \quad - \frac{1}{2} \int_{\Sigma_T} \tilde{h} \cdot \nu [w_{n,xx}^2 + w_{n,yy}^2 + 2\mu w_{n,xx} w_{n,yy} + 2(1 - \mu) w_{n,xy}^2] d\Gamma dt \\ & \quad + \int_{Q_T} f_n \tilde{h} \cdot \nabla w_n d\Omega dt - \frac{1}{2} \int_{Q_T} f_n w_n d\Omega dt + \frac{1}{2} (b(x) w_n, w_n)_{L_2(\Omega)} \Big|_0^T. \end{aligned}$$

Notice that the regularity of the solution given by (2.18) allows us to justify the calculations in [5].

**Step 2: Bounding Linear Terms.** All terms which need to be evaluated at 0 and  $T$ , including the first and second lines and the last term on the right-hand side of (2.33), can be bounded by

$$(2.34) \quad C_1(1 + \gamma^2) E_{w_n,1}(T) + C_2(1 + \gamma^2) E_{w_n,1}(0).$$

Finally, by using duality to split the terms involving  $\alpha_n$  and  $\beta_n$ , noting that all boundary terms involving second derivatives of the solution can be bounded by

second-order traces of the solution and  $L^2(\Omega_T)$ , and taking into account the above estimates, we obtain

$$\begin{aligned}
 \int_0^T E_n(w) dt &\leq C\{(1+\varepsilon^2)E_n(T) + (1+\varepsilon^2)E_n(0) \\
 &\quad + \|f_{1n}\|_{L^2(\Sigma_T)}^2 + \|f_{2n}\|_{L^2(\Sigma_T)}^2 + \|w_n\|_{L^2(\Sigma_T)}^2 \\
 &\quad + \|w_n - \frac{\partial}{\partial \nu} w_n\|_{L^2(\Sigma_T)}^2 + \|f_n - w_n\|_{L^2(\Sigma_T)}^2 + \|\frac{\partial}{\partial \nu} w_n\|_{L^2(\Sigma_T)}^2 + \|w_n\|_{L^2(\Sigma_T)}^2 \\
 &\quad + \|w_n\|_{L^2(\Sigma_T)}^2 + (1+\varepsilon^2)\|\nabla w_n\|_{L^2(\Sigma_T)}^2 \\
 &\quad + \int_0^T \{|\frac{\partial^2}{\partial \nu^2} w_n|^2 + |\frac{\partial^2}{\partial \tau^2} w_n|^2 + |\frac{\partial^2}{\partial \nu \partial \tau} w_n|^2\} d\Gamma dt \\
 &\quad + \int_0^T \int_\Omega h |\nabla w_n|^2 dt + \frac{1}{2} \int_0^T \int_\Omega f_n w_n d\Gamma dt
 \end{aligned}
 \quad (2.35)$$

**Step 6: Bounding Second-Order Traces.** Our next step is to estimate the second-order traces of the function  $w_n$  on the boundary. To accomplish this, it is critical to use the following "regularity" result obtained by microlocal analysis methods.

**Proposition 2.2 ([8], Theorem 2.1).** *Let  $p(t, x)$  be a solution to the following linear problem (in the sense of distributions)*

$$\begin{cases} p_t - \varepsilon^2 \Delta p + \Delta^2 p = F & \text{in } Q_T \\ p(0, \cdot) = p_0; \quad p_t(0, \cdot) = p_1 & \text{in } \Omega \\ \Delta p + (1 - \mu) B_1 p = g_1 & \text{on } \Sigma_T \\ \frac{\partial}{\partial \nu} \Delta p + (1 - \mu) B_2 p - \varepsilon^2 \frac{\partial}{\partial \nu} p_0 = p_2 = g_2 & \text{on } \Sigma_T. \end{cases}
 \quad (2.36)$$

For every  $T > \alpha > 0$  and  $\frac{1}{2} < \varepsilon < 1$ , the following estimate holds:

$$\begin{aligned}
 \int_0^{T-\alpha} \int_\Gamma (|\frac{\partial^2 p}{\partial \tau^2}|^2 + |\frac{\partial^2 p}{\partial \nu^2}|^2 + |\frac{\partial^2 p}{\partial \nu \partial \tau}|^2) d\Gamma dt \\
 \leq C_{T,\alpha} \{ \|F\|_{L^2(\Omega_T)}^2 + \|g_1\|_{L^2(\Sigma_T)}^2 + \|g_2\|_{L^2(\Sigma_T)}^2 + \|p_0\|_{L^2(\Sigma_T)}^2 + \|\frac{\partial}{\partial \nu} p_0\|_{L^2(\Sigma_T)}^2 + \|p_1\|_{L^2(\Omega_T)}^2 + \|p\|_{L^2(\Omega_T)}^2 \}.
 \end{aligned}
 \quad (2.37)$$

Using the result of Proposition 2.2, we shall prove

**Proposition 2.3.** *Let  $w_n$  be the solution to (2.16). Then for any  $T/2 > \alpha > 0$  and  $\frac{1}{2} < \varepsilon < 1$ ,  $w_n$  satisfies the following inequality:*

$$\begin{aligned}
 \int_0^{T-\alpha} \int_\Gamma (|\frac{\partial^2 w_n}{\partial \tau^2}|^2 + |\frac{\partial^2 w_n}{\partial \nu^2}|^2 + |\frac{\partial^2 w_n}{\partial \nu \partial \tau}|^2) d\Gamma dt \\
 \leq C_{T,\alpha,\varepsilon} \{ \|f_n\|_{L^2(\Omega_T)}^2 + \|\alpha_n - \frac{\partial}{\partial \nu} w_n\|_{L^2(\Sigma_T)}^2 + \|f_{1n}\|_{L^2(\Sigma_T)}^2 \\
 + \|\beta_n - w_n\|_{L^2(\Sigma_T)}^2 + \|\frac{\partial^2}{\partial \tau^2} w_n\|_{L^2(\Omega_T)}^2 + \|f_{2n}\|_{L^2(\Sigma_T)}^2 \\
 + \|w_n\|_{L^2(\Sigma_T)}^2 + \|\nabla w_n\|_{L^2(\Sigma_T)}^2 + l.o.(w_n) \}.
 \end{aligned}
 \quad (2.38)$$

**Proof:** We apply the result of Proposition 2.2 to system (2.16) with

$$\begin{aligned}
 F &\equiv f_n \\
 g_1 &\equiv \alpha_n - \frac{\partial}{\partial \nu} w_n - f_{1n} \\
 g_2 &\equiv -\beta_n + w_n - \frac{\partial^2}{\partial \tau^2} w_n + f_{2n}. \quad \square
 \end{aligned}
 \quad (2.39)$$

**Completion of Proof of Lemma 2.2:** Now the result of Lemma 2.2 follows by combining (2.35) applied on  $[\alpha, T - \alpha]$  with Proposition 2.3 and using duality to bound the terms involving  $f_n$ .  $\square$

Recalling Proposition 2.1, we take the limit of (2.31) as  $n \rightarrow \infty$  to obtain a similar inequality for the solution to (1.8), which we state in the following lemma.

### 2.2.2 Estimates for the Original Problem

**Lemma 2.3.** *Let  $(w_0, w_1) \in \mathcal{H}$ . Then the energy of system (1.8) satisfies the following estimate:*

$$(2.40) \quad \begin{aligned} \int_0^T E_w(t) dt &= C_1(1 + \gamma^2)E_w(0) \\ &\leq C_{F, \gamma}(E_w(0), \|b\|_{L_2(0, T; L_2(\Omega))}) \{ \|w_t\|_{L_2(0, T; L_2(\Omega))}^2 \\ &\quad + (1 + \gamma^2) \|\nabla w_t\|_{L_2(0, T; L_2(\Omega))}^2 + l.o.(w) \}. \end{aligned}$$

**Proof:** *Step 1: Approximation Results.* Taking the limit as  $n \rightarrow \infty$  in (2.31), by virtue of (2.11)-(2.15) and Proposition 2.1, we find

$$(2.41) \quad \begin{aligned} \int_0^{T-\alpha} E_w(t) dt &= C_1(1 + \gamma^2)E_{w,1}(T - \alpha) - C_2(1 + \gamma^2)E_{w,1}(\alpha) \\ &\leq C_{F, \gamma, \alpha} \{ \|\hat{f}_1(\frac{\partial}{\partial \nu} w_t)\|_{L_2(0, T; L_2(\Omega))}^2 + \|\hat{g}(w_t) - \frac{\partial}{\partial \tau} \hat{f}_2(\frac{\partial}{\partial \tau} w_t)\|_{L_2(0, T; H^{-1}(\Gamma))}^2 \\ &\quad + \|w_t\|_{L_2(0, T; L_2(\Omega))}^2 + (1 + \gamma^2) \|\nabla w_t\|_{L_2(0, T; L_2(\Omega))}^2 \\ &\quad + \|k(w)\|_{L_2(0, T; L_2(\Omega))}^2 + \|l(w)\|_{L_2(0, T; H^{-1}(\Gamma))}^2 \\ &\quad + \|h(\frac{\partial}{\partial \tau} w)\|_{L_2(0, T; L_2(\Gamma))}^2 + l.o.(w) \}. \end{aligned}$$

*Step 2:* The following bounds can be obtained in a straightforward manner by using Sobolev's Imbeddings and (H-1).

$$(2.42) \quad \|k(w)\|_{L_2(\Omega)} \leq C(\|w\|_{H^{3/2}(\Omega)}) \|w\|_{L_2(\Omega)}$$

$$(2.43) \quad \|l(w)\|_{L_2(\Gamma)} \leq C(\|w\|_{H^{3/2}(\Omega)}) \|w\|_{L_2(\Gamma)}$$

$$(2.44) \quad \|h(\frac{\partial}{\partial \tau} w)\|_{L_2(\Gamma)} \leq C(\|w\|_{H^2(\Omega)}) \|\frac{\partial}{\partial \tau} w\|_{L_2(\Gamma)}$$

Hence by Lemma 2.2,

$$(2.45) \quad \begin{aligned} \|k(w)\|_{L_2(0, T; L_2(\Omega))}^2 + \|l(w)\|_{L_2(0, T; H^{-1}(\Gamma))}^2 + \|h(\frac{\partial}{\partial \tau} w)\|_{L_2(0, T; L_2(\Gamma))}^2 \\ \leq C(E_w(0)) l.o.(w). \end{aligned}$$

*Step 3:* To estimate the first two boundary terms in (2.41), we notice that by virtue of (2.2),

$$(2.46) \quad \|\hat{f}_1(\frac{\partial}{\partial \nu} w_t)\|_{L_2(0, T; L_2(\Omega))} \leq M \|\frac{\partial}{\partial \nu} w_t\|_{L_2(0, T; L_2(\Omega))}^2,$$

$$(2.47) \quad \|\frac{\partial}{\partial \tau} \hat{f}_2(\frac{\partial}{\partial \tau} w_t)\|_{L_2(0, T; H^{-1}(\Gamma))} \leq M \|\frac{\partial}{\partial \tau} w_t\|_{L_2(0, T; L_2(\Gamma))}^2,$$

As for the term with  $\hat{g}(w_t)$ , we use hypothesis (H-3). Indeed, from the imbedding  $H^{-1}(\Gamma) \subset L_1(\Gamma)$ ,

$$\begin{aligned}
 \|\hat{g}(w_t)\|_{L_2(w_t; H^{-1}(\Gamma))}^2 &\leq C \|\hat{g}(w_t)\|_{L_2(w_t; L_1(\Gamma))}^2 \\
 &\leq C \int_0^T \left\{ \int_{\Gamma} |w_t| (|w_t|^r + |b|^r + 1) d\Gamma \right\}^2 dt \\
 (2.48) \quad &\leq C \int_0^T \int_{\Gamma} w_t^2 d\Gamma \left\{ \int_{\Gamma} (|w_t|^r + |b|^r + 1) d\Gamma \right\} dt \\
 &\leq C_T \{ E_w^r(0) + \|b\|_{C^{0,r}(\Gamma; L_2(\Gamma))}^2 + 1 \} \int_{\Sigma_T} w_t^2 d\Gamma dt,
 \end{aligned}$$

where we have used the imbedding  $H^{1/2}(\Gamma) \subset L_{2r}(\Gamma)$  together with the Trace Theorem and the result of Lemma 2.2.

*Step 4:* Noticing that

$$\begin{aligned}
 (2.49) \quad \int_{\alpha}^{T-\alpha} \dot{E}_w(t) dt &= \int_{\alpha}^{T-\alpha} E_w(t) dt \\
 &\quad + \int_{\alpha}^{T-\alpha} \int_{\Gamma} \left\{ \mathcal{L}(w) + \mathcal{H}\left(\frac{\partial}{\partial \tau} w\right) \right\} d\Gamma dt + \int_{\alpha}^{T-\alpha} \int_{\Omega} \mathcal{K}(w) d\Omega dt,
 \end{aligned}$$

where  $\mathcal{K}$ ,  $\mathcal{L}$ ,  $\mathcal{H}$  satisfy (2.42)-(2.44), and applying estimates (2.45)-(2.48) to (2.41) yields

$$\begin{aligned}
 (2.50) \quad \int_{\alpha}^{T-\alpha} E_w(t) dt &\leq C_1 E_w(T-\alpha) + C_2 E_w(\alpha) \\
 &\leq C_{T,\alpha,\gamma} \{ E_w(0), \|b\|_{C^{0,r}(\Gamma; L_2(\Gamma))} \} \left\{ \int_{\Sigma_T} (w_t^2 + |\nabla w_t|^2) d\Gamma dt + l.o.(w) \right\}.
 \end{aligned}$$

From (2.27) and (H-3),

$$(2.51) \quad \int_0^{\alpha} E_w(t) dt + \int_{T-\alpha}^T E_w(t) dt \leq 2\alpha E_w(t)(0).$$

Applying now the result of Lemma 2.2 to the term  $E_w(T-\alpha)$  together with (2.51) yields the conclusion of Lemma 2.3.  $\square$

### 2.2.3 Nonlinear Compactness-Uniqueness Argument

**Lemma 2.4.** *Let  $T > 0$  be sufficiently large. Then*

$$(2.52) \quad l.o.(w) \leq C_T(E_w(0), \|b\|_{C^{0,r}(\Gamma; L_2(\Gamma))}) \int_{\Sigma_T} \{ w_t^2 + |\nabla w_t|^2 + \hat{g}(w_t) w_t \} d\Gamma dt.$$

*Here, the function  $C_T(E_w(0), \|b\|_{C^{0,r}(\Gamma; L_2(\Gamma))})$  does not depend on  $\gamma > 0$ .*

**Proof:** Identical to that in [2].  $\square$

### 2.2.4 Completion of the Proof of Theorem 2.1

Combine the result of Lemma 2.3 with that of Lemma 2.4 and use (2.2) together with  $\|w\|_{L^2(\Omega)} \leq C_1 \|\frac{\partial}{\partial t} w\|_{L^2(\Omega)}$ .  $\square$

## 2.3 Final Estimates: Proof of Theorem 1.2

Denoting  $\mathcal{F} = \int_0^T \{q(w_t)w_t + f_1(\frac{\partial}{\partial t} w_t) \frac{\partial}{\partial t} w_t + f_2(\frac{\partial}{\partial t} w_t) \frac{\partial}{\partial t} w_t\} d\Gamma dt$ , we obtain from Theorem 2.1,

$$(2.53) \quad \int_0^T \dot{E}_w(t) dt - C_1 E_w(T) \leq C_{T,\alpha,\beta}(E_w(0), \|b\|_{C^{0,1}(\Gamma) \times L^2(\Omega)}) \mathcal{F},$$

and by Lemma 2.1,

$$(2.54) \quad \begin{aligned} \int_0^T \dot{E}_w(t) dt &\leq C_{T,\beta}(E_w(0), \|b\|_{C^{0,1}(\Gamma) \times L^2(\Omega)}) \mathcal{F} + C_1 E_w(T) \\ \Rightarrow (T - C_1) E_w(T) &\leq C_{T,\beta}(E_w(0), \|b\|_{C^{0,1}(\Gamma) \times L^2(\Omega)}) \mathcal{F} \\ \Rightarrow \dot{E}_w(T) &\leq C_T(E_w(0), \|b\|_{C^{0,1}(\Gamma) \times L^2(\Omega)}) \mathcal{F}. \end{aligned}$$

Hence, recalling (2.27),

$$(2.55) \quad \frac{\dot{E}_w(T)}{C_T(E_w(0), \|b\|_{C^{0,1}(\Gamma) \times L^2(\Omega)})} \leq \mathcal{F} = E_w(0) - E_w(T).$$

Setting

$$(2.56) \quad p(s) \equiv \frac{s}{C_T(E_w(0), \|b\|_{C^{0,1}(\Gamma) \times L^2(\Omega)})},$$

we have proven the following proposition.

**Proposition 2.4.** *Let  $w$  be the solution to (1.1) and  $E_w(t)$  be the corresponding energy at time  $t$ . If  $T$  is sufficiently large, then there exists a monotone increasing function,  $p$ , such that*

$$(2.57) \quad p(E_w(T)) + \dot{E}_w(T) \leq \dot{E}_w(0).$$

To arrive at the conclusion of Theorem 1.2, we apply a (much more general than needed here) result of Lemma 3.3 in [7].

**Lemma 2.5 ([7], Lemma 3.3).** *Let  $p$  be a positive, increasing function such that  $p(0) = 0$ . Since  $p$  is increasing, we can define a function  $q$  such that  $q(x) = x - (I + p)^{-1}(x)$ . Notice that  $q$  is also an increasing function. Consider a sequence  $s_n$  of positive numbers which satisfy:*

$$(2.58) \quad s_{m+1} + p(s_{m+1}) \leq s_m.$$

Then  $s_m \leq \mathcal{S}(m)$ , where  $\mathcal{S}(t)$  is a solution of a differential equation

$$(2.59) \quad \begin{cases} \frac{d}{dt}\mathcal{S}(t) + q(\mathcal{S}(t)) = 0 \\ \mathcal{S}(0) = s_0. \end{cases}$$

Moreover, if  $p(x) > 0$  for  $x > 0$ , then  $\lim_{t \rightarrow \infty} \mathcal{S}(t) = 0$ .

Applying the result of Proposition 2.4 and noticing that the energy  $\dot{E}_w(t)$  is decreasing, we obtain

$$(2.60) \quad \dot{E}_w(m(T+1)) + p(\dot{E}_w(m(T+1))) \leq \dot{E}_w(mT),$$

for  $m = 0, 1, \dots$ . Thus, applying Lemma 2.5 with

$$(2.61) \quad s_m \equiv \dot{E}_w(mT),$$

yields

$$(2.62) \quad \dot{E}_w(mT) \leq \mathcal{S}(m), \quad m = 0, 1, \dots$$

Setting  $t = mT + \tau$ ,  $0 \leq \tau \leq T$ ,

$$(2.63) \quad E_w(t) \leq E_w(mT) \leq \mathcal{S}(m) \leq \mathcal{S}\left(\frac{t}{T} + \tau\right) \leq \mathcal{S}\left(\frac{t}{T} + 1\right) \text{ for } t \geq T,$$

and noting that in our case since  $q(x)$  is linear,  $\mathcal{S}(t) \sim e^{-\lambda t}$ , completes the proof of Theorem 1.2.  $\square$

### 3. Proof of Theorem 1.3

By arguments similar to those used for the proof of Theorem 2.1 in [2], we obtain the following counterpart of Theorem 2.1.

**Theorem 3.1.** *Let  $w$  be a solution to (1.8) with regularity properties (1.9), (1.10) and let  $T > 0$  be sufficiently large. Then*

$$(3.1) \quad \int_0^T \dot{E}_w(t) dt - C \dot{E}_w(0) \leq C_T(E_w(0)) \left\{ \int_{\Sigma_T} [|w_t|^2 + |\nabla w_t|^2 + g(w_t)w_t] d\Gamma dt \right\}.$$

To proceed with the proof of Theorem 1.3, let the functions  $h(x)$ ,  $h_i(x)$ ,  $i = 1, 2$ , and  $\tilde{h}(x)$  be defined as in (1.14), (1.15). By the hypotheses imposed on functions  $h_i(x)$ , we obtain

$$(3.2) \quad \int_{\Sigma_T} |f_1\left(\frac{\partial}{\partial \nu} w_t\right)|^2 d\Gamma dt = \int_{\Sigma_{A_1}} |f_1\left(\frac{\partial}{\partial \nu} w_t\right)|^2 d\Gamma dt + \int_{\Sigma_{B_1}} |f_1\left(\frac{\partial}{\partial \nu} w_t\right)|^2 d\Gamma dt.$$

where  $\Sigma_{A_1} \triangleq \{(t, x) \in \Sigma_T : |\frac{\partial}{\partial \nu} w_t| \leq 1\}$  and  $\Sigma_{B_1} = \Sigma_T \setminus \Sigma_{A_1}$ . Hence, using hypothesis (H-2) on  $\Sigma_{B_1}$ , we find

$$\begin{aligned}
 (3.3) \quad & \int_{\Sigma_T} |\frac{\partial}{\partial \nu} w_t|^2 d\Gamma dt + \int_{\Sigma_T} |f_1(\frac{\partial}{\partial \nu} w_t)|^2 d\Gamma dt \\
 & \leq \int_{\Sigma_{A_1}} |\frac{\partial}{\partial \nu} w_t|^2 + |f_1(\frac{\partial}{\partial \nu} w_t)|^2 d\Gamma dt \\
 & \quad + (M + \frac{1}{m}) \int_{\Sigma_{B_1}} f_1(\frac{\partial}{\partial \nu} w_t) \frac{\partial}{\partial \nu} w_t d\Gamma dt \\
 & \leq \int_{\Sigma_{A_1}} h_1(\frac{\partial}{\partial \nu} w_t) f_1(\frac{\partial}{\partial \nu} w_t) d\Gamma dt \\
 & \quad + (M + \frac{1}{m}) \int_{\Sigma_T} f_1(\frac{\partial}{\partial \nu} w_t) \frac{\partial}{\partial \nu} w_t d\Gamma dt.
 \end{aligned}$$

Similarly, the same argument applied to  $f_2$  yields

$$\begin{aligned}
 (3.4) \quad & \int_{\Sigma_T} |\frac{\partial}{\partial \tau} w_t|^2 d\Gamma dt + \int_{\Sigma_T} |f_2(\frac{\partial}{\partial \tau} w_t)|^2 d\Gamma dt \\
 & \leq \int_{\Sigma_{A_2}} h_2(\frac{\partial}{\partial \tau} w_t) f_2(\frac{\partial}{\partial \tau} w_t) d\Gamma dt + (M + \frac{1}{m}) \int_{\Sigma_T} f_2(\frac{\partial}{\partial \tau} w_t) \frac{\partial}{\partial \tau} w_t d\Gamma dt.
 \end{aligned}$$

Recall

$$(3.5) \quad \tilde{h}_i(x) \equiv h_i(\frac{x}{\max \Sigma_T}).$$

Then, by Jensen's inequality,

$$\begin{aligned}
 (3.6) \quad & \int_{\Sigma_T} \{ |w_t|^2 + |\nabla w_t|^2 + g(w_t) w_t \} d\Gamma dt \\
 & \leq C_1 \int_{\Sigma_T} \{ g(w_t) w_t + f_1(\frac{\partial}{\partial \nu} w_t) \frac{\partial}{\partial \nu} w_t + f_2(\frac{\partial}{\partial \tau} w_t) \frac{\partial}{\partial \tau} w_t \} d\Gamma dt \\
 & \quad + C_2 \left[ \tilde{h}_1(\int_{\Sigma_T} f_1(\frac{\partial}{\partial \nu} w_t) \frac{\partial}{\partial \nu} w_t d\Gamma dt) + \tilde{h}_2(\int_{\Sigma_T} f_2(\frac{\partial}{\partial \tau} w_t) \frac{\partial}{\partial \tau} w_t d\Gamma dt) \right] \\
 & \leq C_1 \int_{\Sigma_T} \{ g(w_t) w_t + f_1(\frac{\partial}{\partial \nu} w_t) \frac{\partial}{\partial \nu} w_t + f_2(\frac{\partial}{\partial \tau} w_t) \frac{\partial}{\partial \tau} w_t \} d\Gamma dt \\
 & \quad + C_2 \sum_{i=1}^2 \tilde{h}_i(\int_{\Sigma_T} \{ g(w_t) w_t + f_1(\frac{\partial}{\partial \nu} w_t) \frac{\partial}{\partial \nu} w_t + f_2(\frac{\partial}{\partial \tau} w_t) \frac{\partial}{\partial \tau} w_t \} d\Gamma dt),
 \end{aligned}$$

where the last inequality follows from the monotonicity of the functions  $\tilde{h}_i$ .

Denoting  $\mathcal{F} \equiv \int_{\Sigma_T} \{ g(w_t) w_t + f_1(\frac{\partial}{\partial \nu} w_t) \frac{\partial}{\partial \nu} w_t + f_2(\frac{\partial}{\partial \tau} w_t) \frac{\partial}{\partial \tau} w_t \} d\Gamma dt$ , we obtain from Theorem 3.1 and (3.6),

$$(3.7) \quad \int_{\alpha}^{T-\alpha} E_w(t) dt - C_1 E_w(0) \leq C_{T,\alpha,\alpha}(E_w(0)) [\mathcal{F} + \tilde{h}(\mathcal{F})].$$

Since

$$(3.8) \quad \int_0^{\alpha} E_w(t) dt + \int_{T-\alpha}^T E_w(t) dt \leq 2\alpha E_w(0),$$

we find

$$(3.9) \quad \int_0^T E_w(t) dt - C_{1,\alpha} E_w(0) \leq C_{T,\alpha,\alpha}(E_w(0)) [\mathcal{F} + \tilde{h}\mathcal{F}],$$



and by Lemma 2.1,

$$\begin{aligned}
 (3.10) \quad & \int_0^T E_w(t) dt \leq C_{T,\alpha,\beta}(E_w(0))[\mathcal{F} + \hat{h}\mathcal{F}] + C_{1,\alpha}E_w(0) \\
 & \Rightarrow (T - C_{1,\alpha})E_w(T) \leq C_{T,\alpha,\beta}(E_w(0))[\mathcal{F} + \hat{h}(\mathcal{F})] \\
 & \Rightarrow E_w(T) \leq C_T(E_w(0))[\mathcal{F} + \hat{h}(\mathcal{F})].
 \end{aligned}$$

Hence, recalling (2.27),

$$(3.11) \quad (I + \hat{h})^{-1} \left( \frac{E_w(T)}{C_T(E_w(0))} \right) \leq \mathcal{F} - E_w(0) - E_w(T).$$

Setting

$$(3.12) \quad p(s) \equiv (I + \hat{h})^{-1} \left( \frac{s}{C_T(E_w(0))} \right),$$

we have the result of Proposition 2.4 valid with the above function  $p(s)$ .

As in section 2.3, applying the result of Proposition 2.4, we obtain

$$(3.13) \quad E_w(m(T+1)) + p(E_w(m(T+1))) \leq E_w(mT),$$

for  $m = 0, 1, \dots$ . Thus, applying Lemma 2.5 with

$$(3.14) \quad s_m \equiv E_w(mT),$$

yields

$$(3.15) \quad E_w(mT) \leq \mathcal{S}(m), \quad m = 0, 1, \dots$$

Setting  $t = mT + \tau$ ,  $0 \leq \tau < T$ ,

$$(3.16) \quad E_w(t) \leq E_w(mT) \leq \mathcal{S}(m) \leq \mathcal{S}\left(\frac{t-\tau}{T}\right) \leq \mathcal{S}\left(\frac{t}{T}\right) - 1 \text{ for } t \leq T,$$

which completes the proof of Theorem 1.3.  $\square$

#### 4. Proof of Theorem 1.1

The first part of Theorem 1.1 (existence) follows by applying the result of the main theorem in [6] within the framework described in the first section of [1] where a system of coupled plate equations is considered. The details are omitted for lack of space. We shall concentrate on the second part of Theorem 1.1, i.e., estimate (1.7).

#### 4.1 Change of Variables

We introduce a change of variables,  $w \equiv v - u$ . Then  $w$  satisfies

$$(4.1) \quad \begin{cases} w_{tt} - \gamma^2 \Delta w_{tt} + \Delta^2 w = k_1(\cdot w) + k_2(w) \\ w(0, \cdot) = v_0 - u_0; \quad w_t(0, \cdot) = v_1 - u_1 \\ \Delta w + (1 - \mu) \mathcal{B}_1 w = -[f_1(\frac{\partial}{\partial \nu} w_t + \frac{\partial}{\partial \tau} u_t) - f_1(\frac{\partial}{\partial \nu} u_t)] \\ \frac{\partial}{\partial \nu} \Delta w + (1 - \mu) \mathcal{B}_2 w - \gamma^2 \frac{\partial}{\partial \nu} w_{tt} - w + \frac{\partial}{\partial \tau} h_1(\frac{\partial}{\partial \tau} w) - \frac{\partial}{\partial \tau} h_1(-\frac{\partial}{\partial \tau} w) \\ \quad - l_1(w) + l_2(\cdot w) \\ \quad = g(w_t + u_t) - g(u_t) - \frac{\partial}{\partial \tau} f_2(\frac{\partial}{\partial \tau} w_t + \frac{\partial}{\partial \tau} u_t) + \frac{\partial}{\partial \tau} f_2(\frac{\partial}{\partial \tau} u_t). \end{cases}$$

By virtue of (4.6),

$$(4.2) \quad \frac{\partial}{\partial \nu} u_t, \quad \frac{\partial}{\partial \tau} u_t \in L_2(\Sigma_\infty).$$

Moreover,

$$(4.3) \quad \|u_t\|_{C^{1,0} \times H^1(\Omega)} \leq C(E_u(0), E_v(0)).$$

Hence, by Sobolev's Imbeddings, for any  $r \geq 1$ ,

$$(4.4) \quad \|u_t\|_{C^{1,0} \times L_{2r}(\Omega)} \leq C(E_u(0), E_v(0)).$$

Setting  $a_1 \equiv \frac{\partial}{\partial \nu} u_t$ ,  $a_2 \equiv \frac{\partial}{\partial \tau} u_t$ ,  $b \equiv u_t$ . We conclude that  $a_i \in L_2(\Sigma_\infty)$  and  $b \in C(0, \infty; L_{2r}(\Gamma))$ . On the other hand, functions defined by

$$\begin{aligned} k(s) &\equiv k_2(s) + k_1(-s) \\ l(s) &\equiv l_1(s) + l_2(-s) \\ h(s) &\equiv h_1(s) + h_2(-s) \end{aligned}$$

comply with hypothesis (H-1).

Thus we are in a position to apply the result of Theorem 4.2 to equation (4.1). This yields

**Theorem 4.1.** *Let  $w(t)$  be any solution of finite energy corresponding to system (4.1). Then there exist constants  $C, \omega > 0$  such that*

$$(4.5) \quad \|(u - v)(t)\|_{H^2(\Omega)} + \|(u_t - v_t)(t)\|_{H^1(\Omega)} \leq C e^{-\omega t},$$

where  $C, \omega$  may depend on  $\|u_0\|_{H^2(\Omega)}, \|v_0\|_{H^2(\Omega)}, \|u_1\|_{H^1(\Omega)}, \|v_1\|_{H^1(\Omega)}$ .

## 4.2 Analysis of "u" Equation

Using the variable  $w$ , we see that the equation for  $u$  (see (1.1)) is equivalent to

$$(4.6) \quad \begin{cases} u_{tt} - \gamma^2 \Delta u_{tt} + \Delta^2 u = k_2(w) & \text{in } Q_\infty \\ u(0, \cdot) = u_0; \quad u_t(0, \cdot) = u_1 & \text{in } \Omega \\ \Delta u + (1 - \mu) \mathcal{B}_1 u = -f_1(\frac{\partial}{\partial \nu} u_t) & \text{on } \Sigma_\infty \\ \frac{\partial}{\partial \nu} \Delta u + (1 - \mu) \mathcal{B}_2 u - \gamma^2 \frac{\partial}{\partial \nu} u_{tt} - u \\ \quad = g(u_t) - \frac{\partial}{\partial \tau} f_2(\frac{\partial}{\partial \tau} u_t) + l_2(-w) - \frac{\partial}{\partial \tau} h_2(\frac{\partial}{\partial \tau} (-w)) & \text{on } \Sigma_\infty, \end{cases}$$

where from Theorem 4.1,  $w(t)$  satisfies

$$(4.7) \quad \|w(t)\|_{H^2(\Omega)} \leq C'e^{-\omega t},$$

with the constants  $C$  and  $\omega$  depending on  $E_v(0)$ ,  $E_u(0)$ ,  $\|u_0\|_{H^2(\Omega)}$ ,  $\|v_0\|_{H^2(\Omega)}$ ,  $\|u_1\|_{H^1(\Omega)}$ ,  $\|v_1\|_{H^1(\Omega)}$ .

**Proposition 4.1.** *Let  $E_v(0) \leq R$ ,  $E_u(0) \leq R$ . Then there exist constants  $C$ ,  $\omega > 0$  such that for all  $t \geq 0$ ,*

$$(4.8) \quad \|k_2(w)(t)\|_{L_2(\Omega)} \leq C'e^{-\omega t}$$

$$(4.9) \quad \|l_2(w)(t)\|_{L_2(\Gamma)} \leq C'e^{-\omega t}$$

$$(4.10) \quad \left\| \frac{\partial}{\partial \tau} h_2\left(\frac{\partial}{\partial \tau}(-w(t))\right) \right\|_{H^{-1}(\Gamma)} \leq C'e^{-\omega t}.$$

**Proof:** Inequalities (4.8) and (4.9) follow direct from (4.7), hypothesis (H-1) and the imbedding  $H^2(\Omega) \subset C(\Omega)$ . As for (4.10), we have

$$(4.11) \quad \begin{aligned} \left\| \frac{\partial}{\partial \tau} h_2\left(\frac{\partial}{\partial \tau}(-w(t))\right) \right\|_{H^{-1}(\Gamma)}^2 &\leq C' \|h_2\left(\frac{\partial}{\partial \tau}(-w(t))\right)\|_{L_2(\Gamma)}^2 \\ &\leq C' \int_\Gamma \left\{ \left| \frac{\partial}{\partial \tau} w(t) \right|^{2k} + \left| \frac{\partial}{\partial \tau} w(t) \right|^2 \right\} d\Gamma \\ &\leq C' \{ \|w(t)\|_{H^2(\Omega)}^2 + \|w(t)\|_{H^2(\Omega)}^{2k} \}, \end{aligned}$$

where the second inequality follows from the imbedding  $H^{1/2}(\Gamma) \subset L_{2k}(\Gamma)$  followed by the Trace Theorem. (4.7) together with (4.11) implies (4.10).  $\square$

We next consider the following nonautonomous linear problem.

$$(4.12) \quad \begin{cases} u_{tt} - \gamma^2 \Delta w_{tt} + \Delta^2 u = k(t, x) & \text{in } Q_\infty \\ u(0, \cdot) = u_0; \quad u_t(0, \cdot) = u_1 & \text{in } \Omega \\ \Delta u + (1 - \mu) \mathcal{B}_1 u = -f_1(t, x) \frac{\partial}{\partial \nu} u_t & \text{on } \Sigma_\infty \\ \frac{\partial}{\partial \nu} \Delta u + (1 - \mu) \mathcal{B}_2 u - \gamma^2 \frac{\partial}{\partial \nu} u_{tt} - u \\ \quad = g(t, x) u_t - \frac{\partial}{\partial \tau} f_2(t, x) \frac{\partial}{\partial \tau} u_t + l(t, x) & \text{on } \Sigma_\infty, \end{cases}$$

where

$$(4.13) \quad \|k(t)\|_{L_2(\Omega)} \leq C_0 e^{-\omega_0 t}$$

$$(4.14) \quad \|l(t)\|_{H^{-1}(\Omega)} \leq C_0 e^{-\omega_0 t}$$

$$(4.15) \quad m \leq f_i(t, x) \leq M \quad (t, x) \in \Sigma_\infty$$

$$(4.16) \quad g(t, x) \geq 0 \quad (t, x) \in \Sigma_\infty$$

$$(4.17) \quad \|g(t, \cdot)\|_{L_2(\Gamma)} \leq M$$

Notice that a homogeneous system (4.12) with  $k \equiv l \equiv 0$  is exponentially stable. Hence, one can show by using *linear* evolution methods, that this stability is preserved for a system with nonhomogeneous, but exponentially decaying terms. This is stated below.

**Lemma 4.1.** *Let  $u$  be a solution to (4.12) subject to assumptions (4.13)-(4.17). Then there exist constants  $C, \omega > 0$  depending on  $C_0, \omega_0, M, m$  such that for all  $t \geq 0$ ,*

$$(4.18) \quad \|u(t)\|_{H^2(\Omega)} + \|u_t(t)\|_{H^1(\Omega)} \leq C e^{-\omega t} (\|u_0\|_{H^2(\Omega)} + \|u_1\|_{H^1(\Omega)} + 1).$$

We note that the solution  $u$  to (4.7) satisfies (4.12) with

$$(4.19) \quad k(t, x) \equiv k_2(w(t, x)), \quad l(t, x) \equiv l_2(-w(t, x)) - \frac{\partial}{\partial \tau} l_2(-\frac{\partial}{\partial \tau} w(t, x))$$

$$(4.20) \quad f_1(t, x) \equiv \frac{f_1(\frac{\partial}{\partial \tau} u_t(t, x))}{\frac{\partial}{\partial \tau} u_t(t, x)}$$

$$(4.21) \quad f_2(t, x) \equiv \frac{f_2(\frac{\partial}{\partial \tau} u_t(t, x))}{\frac{\partial}{\partial \tau} u_t(t, x)}$$

$$(4.22) \quad g(t, x) \equiv \frac{g(u_t(t, x))}{u_t(t, x)}.$$

Hypotheses assumed in (H-1) together with the result of Proposition 4.1 imply that (4.13)-(4.17) hold true.

Thus we are in a position to apply the result of Lemma 4.1 to the solution  $u$  of equation (4.1). This yields

**Theorem 4.2.** *Let  $u$  be a solution to (4.5). Then there exist constants  $C, \omega > 0$  depending on  $E_v(0), E_u(0)$  such that*

$$(4.23) \quad \|u(t)\|_{H^2(\Omega)} + \|u_t(t)\|_{H^1(\Omega)} \leq C e^{-\omega t}.$$

Combining together the results of Theorem 4.1 and Theorem 4.2 and recalling  $v \equiv w + u$  yields the final conclusion, (1.7) in Theorem 1.1.  $\square$

## References

1. A. Favini and I. Lasiecka. Well-posedness and regularity of second-order abstract equations arising in hyperbolic-like problems with nonlinear boundary conditions. *Osaka Journal of Mathematics*. To appear.
2. M.A. Horn and I. Lasiecka. Global stabilization of a dynamic von Kármán plate with nonlinear boundary feedback. *Applied Mathematics and Optimization*. To appear.
3. V. Komornik. Rapid boundary stabilization of the wave equation. *SIAM Journal of Control and Optimization*, 29 (1):197-208 January 1991.
4. V. Komornik and E. Zuazua. A direct method for the boundary stabilization of the wave equation. *J. Math. Pures et. Appl.*, 69:33-54, 1990.
5. J. E. Lagnese. *Boundary Stabilization of Thin Plates*. Society for Industrial and Applied Mathematics, Philadelphia, 1989.
6. I. Lasiecka. Existence and uniqueness of the solutions to second order abstract equations with nonlinear and nonmonotone boundary conditions. *Journal of Nonlinear Analysis, Methods and Applications*. To appear.
7. I. Lasiecka and D. Tataru. Uniform boundary stabilization of semilinear wave equations with nonlinear boundary damping. *Differential and Integral Equations*, 6(3):507-533, 1993.
8. I. Lasiecka and R. Triggiani. Sharp trace estimates of solutions to Kirchhoff and Euler-Bernoulli equations. *Applied Mathematics and Optimization*, 28:277-306, 1993.
9. W. Littman. Boundary control theory for beams and plates. In *Proceedings of 24th Conference on Decision and Control*, pages 2007-2009, 1985.
10. M. Najafi, G. Sarhangi, and H. Wang. The study of the stabilizability of the coupled wave equations under various end conditions. In *Proceedings of the 31st IEEE Conference on Decision and Control*, Tucson, Arizona, 1992.
11. M. Najafi and H. Wang. Exponential stability of wave equations coupled in parallel by viscous damping. In *Proceedings of the 32nd IEEE Conference on Decision and Control*, San Antonio, Texas, 1993.
12. B. Rao. Stabilization of a Kirchhoff plate equation in star-shaped domain by nonlinear boundary feedback. *Nonlinear Analysis*, 20(6):605-626, 1993.

Mary Ann Horn  
 School of Mathematics  
 University of Minnesota  
 Minneapolis, Minnesota 55455, USA

Irena Lasiecka  
 Department of Applied Mathematics  
 University of Virginia  
 Charlottesville, Virginia 22903, USA

## BOUNDARY TEMPERATURE CONTROL FOR THERMALLY COUPLED NAVIER-STOKES EQUATIONS

KAZUFUMI ITO

Center for Research in Scientific Computation  
Department of Mathematics  
North Carolina State University

**ABSTRACT.** In this paper the optimal control problem for the thermally coupled incompressible Navier-Stokes equations by the Dirichlet boundary temperature control is discussed. Well-posedness and existence of the optimal control for the finite-time horizon problem and optimal control problem for the stationary equations are established. Necessary optimality conditions are also obtained.

1991 *Mathematics Subject Classification.* 76D05, 93C20, 49B22

*Key words and phrases.* Boussinesq equation, boundary temperature control, necessary optimality condition.

### 1. Introduction

In this paper we discuss the optimal control problem of the thermally coupled incompressible Navier-Stokes equations. Consider the following optimal control problem

$$(1.1) \quad \begin{aligned} &\text{minimize} \quad J(g) = \int_0^T \left( \varphi(u(t), T(t) - T_0) + \frac{\beta}{2} \|g(t) - T|_{\partial\Omega}(t)\|^2 \right) dt \\ &\text{over } g(t) \in \mathcal{C} \end{aligned}$$

subject to

$$(1.2) \quad \begin{aligned} &\frac{\partial}{\partial t} u + u \cdot \nabla u + \nabla p = \nu \Delta u + \gamma (T - T_0) e_d + f \\ &\nabla \cdot u = 0, \quad u|_{\Gamma} = 0, \end{aligned}$$

$$\frac{\partial}{\partial t} T + u \cdot \nabla T = \nabla \cdot (\kappa \nabla T), \quad T = g \text{ on } \Gamma$$

where  $f \in L^2(\Omega)^d$  is a source field,  $u$ ,  $p$ ,  $T$  stand for the nondimensionalized velocity vector in  $R^d$  with  $d = 2, 3$ , pressure, and temperature, respectively and

$\mathcal{C}$  is the closed convex set in  $L^2(\Gamma)$  such that

$$(1.3) \quad \bar{T}_1 \leq g(x) \leq \bar{T}_2 \quad \text{a.e. } x \in \Gamma$$

Here,  $\bar{T}_1 \leq T_0 \leq \bar{T}_2$  and  $\gamma = \frac{\bar{g}}{T_0}$  where  $T_0 > 0$  is a constant reference temperature and  $\bar{g}$  is the gravitational constant, the vector  $e_d$  denotes the  $d$ -th unit vector of  $R^d$ . Throughout this paper we assume that  $\Omega$  is sufficiently smooth,  $\nu, \kappa$  are positive constants and  $\theta(t, \cdot) = T(t, \cdot) - T_0$ . This control problem and the corresponding problem for the stationary flow (see, section 2) are motivated from control of the transport process in the high pressure vapor transport (HPVT) reactor [IST1]. For example we consider the Scholz geometry depicted in Figure 1. The source material and the growing crystal are sealed in a fused silica ampoule that is heated by a furnace liner at its outer cylindrical surface. The substrate  $\Gamma_0$  (the single crystal) is located on a fused silica window (W) which is cooled by a jet of helium gas from the outer surface. HPVT processes are based on physical vapor transport and can be described very roughly as proceeding via evaporation at the polycrystalline source and condensation at the surface of the cooler substrate. The system of equations (1.2) is called the Boussinesq equation where we assume that the flow is incompressible and the transport phenomena of a single (carrier) gas is modeled. The objective of our control problem includes the uniformity of the temperature in a neighborhood of the substrate (W). For example, the performance index  $\varphi$  appearing (1.1) is given by

$$\varphi(u(t), T(t) - T_0) = \frac{1}{2} \int_{\Omega} (|u(t, x) - u_d(x)|^2 + |T(t, x) - T_d(x)|^2) dx,$$

where the pair  $(u_d(x), T_d(x))$  is the desired state. We refer [IST1], [IST2] for the specific choice of  $(u_d, T_d)$ .

Also we consider a family of control problems parameterized by  $\epsilon > 0$ :

$$(1.4) \quad \text{minimize } J(g) \quad \text{over } g(t) \in \mathcal{C}$$

subject to

$$(1.5) \quad \begin{aligned} \frac{\partial}{\partial t} u + u \cdot \nabla u + \nabla p &= \nu \Delta u + \gamma (T - T_0) e_d + f \\ \nabla \cdot u &= 0, \quad u|_{\Gamma} = 0, \\ \frac{\partial}{\partial t} T + u \cdot \nabla T &= \nabla \cdot (\kappa \nabla T), \end{aligned}$$

$$\kappa n \cdot \nabla T(t, \cdot) = \frac{1}{\epsilon} (g(t, \cdot) - T(t, \cdot)) \text{ on } \Gamma,$$

where  $n$  is the outward normal vector at the boundary  $\Gamma$ . The boundary condition for  $T$  is given by the Newton's law of cooling. Problem (1.4) - (1.5) is discussed in [IST2] and is much nicer to be dealt with both in the theoretical and numerical

point of view since it is naturally formulated under the Gelfand-triple framework in [LM],[IST2]. In this paper we develop a weak formulation of the Dirichlet boundary control problem in the thermal equation of (1.2) (see, Definitions 2.1 and 3.1) which is based on the transposition of the Gelfand triple. A solution to equation (1.2) is defined by an asymptotic limit of solutions to equation (1.5) when  $\epsilon \rightarrow 0^+$ . This limiting procedure is necessary since our weak formulation does not allow us to use the (weak) maximum principle to obtain *a priori* bound of solutions. Thus we analyze problem (1.1)–(1.2) as the limiting case of problem (1.1) and (1.5).

The outline of the paper is as follows. In section 2 we discuss the corresponding control problems (2.1)–(2.2) for the stationary flow. Basic functional space framework and the definition of weak solutions to the steady-state equation are given. The existence of weak solutions then is shown by the limiting procedure described above where *a priori*  $L^\infty$ -bound of the thermal component  $\theta$  is used in an essential way. Necessary optimality condition is obtained by applying the Lagrange multiplier theory [MZ]. In section 3 we discuss the finite time horizon problem (1.1)–(1.2). The existence of weak solutions to (1.2) and necessary optimality condition for  $d = 2$  are established by the limiting procedure.

## 2. Stationary Case

In this section we consider control problems for the stationary flow:

$$(2.1) \quad \text{minimize } J(g) = \varphi(u, T - T_0) + \frac{\beta}{2} \|g - T_0\|_{L^2(\Gamma)}^2 \quad \text{over } g \in C$$

subject to

$$-\nu \Delta u + u \cdot \nabla u + \nabla p = \gamma (T - T_0) e_d + f$$

$$(2.2) \quad \nabla \cdot u = 0, \quad u|_\Gamma = 0,$$

$$-\kappa \Delta T + u \cdot \nabla T = 0, \quad T = g \quad \text{on } \Gamma.$$

**2.1. Wellposedness.** In this section we discuss existence and regularity of solutions to (2.2). Let  $V_0$  be the divergence free subspace of  $(H_0^1(\Omega))^d$  [GR] and  $H_0$  is defined by

$$H_0 = \{\phi \in L^2(\Omega)^d : \nabla \cdot \phi = 0 \text{ and } n \cdot \phi = 0 \text{ on } \Gamma\}.$$

$H_0$  is equipped with the natural  $L^2$ -norm and  $V_0$  is equipped with  $\|u\|_{V_0}^2 = \|\nabla u\|_{L^2(\Omega)}^2$ . Let  $H_1 = L^2(\Omega)$  equipped with the natural  $L^2$ -norm and  $V_1 = H_0^1(\Omega)$  equipped with  $\|\phi\|_{V_1}^2 = \|\nabla \phi\|_{L^2(\Omega)}^2$ . If we identify  $H_1$  with  $H_1^*$  then  $V_1 \subset H_1 = H_1^* \subset V_1^*$ . Let  $\Delta$  denote the Laplacian and  $\Delta_0 = \Delta$  with  $\text{dom}(\Delta_0) = H^2(\Omega) \cap H_0^1(\Omega)$ . Then  $-\Delta_0$  is positive selfadjoint operator on  $H_1$ , and  $-\Delta_0 \in \mathcal{L}(V_1, V_1^*)$  with

$$\langle -\Delta_0 \phi, \psi \rangle = (\nabla \phi, \nabla \psi)_{L^2} \quad \text{for } \phi, \psi \in V_1$$



where  $V_1^* = H^{-1}(\Omega)$  and  $\langle \cdot, \cdot \rangle$  denotes the dual product of  $V_1^* \times V_1$ . Moreover,  $V_1^*$  can be equipped with

$$(2.3) \quad \|\phi\|_{V_1^*}^2 = \langle \phi, (-\Delta_0)^{-1} \phi \rangle,$$

i.e.,  $(-\Delta_0)^{-1} \in \mathcal{L}(V_1^*, V_1)$  is the Riesz map of  $V_1^*$ . Consider the second equation of (2.2) when  $u \in V_0$  is fixed. By Green's formula we have from the second equation of (2.2)

$$(u \cdot \nabla \theta - \kappa \Delta \theta, \psi) = \kappa (\theta, \phi) + \kappa (g - T_0, n \cdot \nabla \psi)_\Gamma - (\theta, u \cdot \nabla \psi) = 0$$

for  $\psi = (-\Delta_0)^{-1} \phi$ ,  $\phi \in H_1$ , where we assumed that  $\theta = T - T_0$  is sufficiently smooth and  $(f, g)_\Gamma = \int_\Gamma f(s)g(s)ds$ . Thus, the second equation of (2.2) can be written as

$$(2.4) \quad \kappa (\theta, \phi)_{L^2} - (\theta, u \cdot \nabla \psi)_{L^2} + \kappa (g - T_0, n \cdot \nabla \psi)_\Gamma = 0, \quad \text{where } \psi = (-\Delta_0)^{-1} \phi$$

for  $\phi \in H_1 = L^2(\Omega)$ . Here, we assume that the domain  $\Omega$  is sufficiently smooth so that

$$\|(-\Delta_0)^{-1} \phi\|_{H^2(\Omega)} \leq \alpha \|\phi\|_{L^2(\Omega)}, \quad \phi \in L^2(\Omega)$$

for some  $\alpha > 0$ . Hence, we have  $\nabla \psi \in H^1(\Omega)$  and  $n \cdot \nabla \psi \in H^{\frac{1}{2}}(\Gamma)$  for  $\psi = (-\Delta_0)^{-1} \phi$ ,  $\phi \in L^2(\Omega)$ . Let  $W$  be the completion of  $L^2(\Omega)$  with respect to norm  $\|(-\Delta_0)^{-1} \phi\|_{L^2}$ . Then,  $W = (H_1)^*$  where  $V_1^*$  is taken to be the pivoting space equipped with inner product (2.3). The dual product on  $W \times H_1$  is defined as

$$\langle \theta, \phi \rangle_{W \times H_1} = \langle (-\Delta_0)^{-1} \theta, \phi \rangle, \quad \theta \in W \text{ and } \phi \in H_1.$$

Thus,  $H_1 \subset V_1^* = (V_1^*)^* \subset W$  defines the Gelfand triple and results in the transposition of the standard triple:  $H_0^1(\Omega) \subset L^2(\Omega) = L^2(\Omega)^* \subset H^{-1}(\Omega)$  by one Sobolev index. Moreover, (2.4) is equivalently written as

$$-\kappa \Delta \theta + \nabla \cdot (u\theta) + r = 0 \quad \text{in } W$$

where  $r \in W$  is defined by

$$(r, \phi)_{W \times H_1} = \kappa (g - T_0, n \cdot \nabla \psi)_\Gamma, \quad \text{with } \psi = (-\Delta_0)^{-1} \phi, \phi \in H_1.$$

Now, we define the weak solution to (2.2) by

**Definition 2.1.** The pair  $(u, \theta) \in V_0 \times L^2(\Omega)$  is a weak solution to (2.2) if

$$(2.5a) \quad \nu (\nabla u, \nabla v) + b(u, u, v) = (\gamma \theta e_d + f, v)$$

for all  $v \in V_0$  and

$$(2.5b) \quad \kappa (\theta, \phi)_{L^2} - (\theta, u \cdot \nabla \psi)_{L^2} + \kappa (g - T_0, n \cdot \nabla \psi)_\Gamma = 0, \quad \text{where } \psi = (-\Delta_0)^{-1} \phi$$

for all  $\phi \in L^2(\Omega)$ , where the trilinear form  $b$  on  $(V_0)^3$  is defined by

$$b(u, v, w) = \int_\Omega (u \cdot \nabla v) \cdot w \, dx \quad \text{for } u, v, w \in V_0.$$

It follows from [Te1] that

**Lemma 2.2.** *The trilinear form  $b$  satisfies*

- (a)  $|b(u, v, w)| \leq |u|_{L^4} |\nabla v|_{V_0} |w|_{L^4} \leq M_1 |u|_{V_0} |v|_{V_0} |w|_{V_0}$
- (b)  $b(u, v, w) + b(u, w, v) = 0$  in particular  $b(u, v, v) = 0$
- (c)  $|b(u, v, w)| \leq M_2 |u|_{H_0}^{1/2} |u|_{V_0}^{1/2} |v|_{H_0}^{1/2} |v|_{V_0}^{1/2} |w|_{V_0}$  for  $d = 2$
- (d)  $|b(u, v, w)| \leq M_2 |u|_{H_0}^{1/4} |u|_{V_0}^{3/4} |v|_{H_0}^{1/4} |v|_{V_0}^{3/4} |w|_{V_0}$  for  $d = 3$

for  $u, v, w \in V_0$ .

The sesquilinear form on  $H_1 \times H_1$  defined by

$$\kappa(\phi_1, \phi_2)_{L^2} - (\phi_1, u \cdot \nabla \psi)_{L^2} \quad \text{where } \psi = (-\Delta_0)^{-1} \phi_2$$

for  $\phi_i, i = 1, 2 \in H_1$  and  $u \in V_0$  given, is bounded but is not necessarily  $H_1$ -coercive unless  $|u|_{V_0}$  is sufficiently small. Thus, one cannot show that (2.5) has a solution  $(u, \theta) \in V_0 \times H_1$  in general. We overcome this difficulty by an approach in [It] in which a solution to (2.5) is defined by a weak limit of  $(u_\epsilon, \theta_\epsilon)$  where  $(u_\epsilon, \theta_\epsilon) \in V_0 \times H^1(\Omega)$  is a weak solution to

$$\begin{aligned} (2.6) \quad & -\nu \Delta u_\epsilon + u_\epsilon \cdot \nabla u_\epsilon + \nabla p = \gamma \theta_\epsilon e_d + f \\ & -\kappa \Delta \theta_\epsilon + u_\epsilon \cdot \nabla \theta_\epsilon = 0, \\ & \kappa n \cdot \theta_\epsilon = \frac{1}{\epsilon} (g - T_0 - \theta_\epsilon) \text{ on } \Gamma. \end{aligned}$$

A variational form of (2.6) is given by (2.5a) and

$$(2.7) \quad \kappa(\nabla \theta, \nabla \psi)_{L^2} + (u \cdot \nabla \theta, \psi) + \frac{1}{\epsilon} (\theta, \psi)_\Gamma = \frac{1}{\epsilon} (g - T_0, \psi)_\Gamma$$

for  $\psi \in H^1(\Omega)$ . Then we have

**Lemma 2.3.** *For  $\epsilon > 0$  and  $g \in L^2(\Gamma)$  equation (2.5a) and (2.7) has a solution  $(u_\epsilon, \theta_\epsilon)$  in  $V_0 \times H^1(\Omega)$ . Moreover if  $T_1 \leq g \leq T_2$  a.e. in  $\Gamma$  then  $T_1 - T_0 \leq \theta_\epsilon \leq T_2 - T_0$  a.e. in  $\Omega$ .*

**Proof:** Let  $\bar{u} \in V_0$  be given. Note that  $(\bar{u} \cdot \nabla \psi, \psi) = -\frac{1}{2} (\nabla \cdot \bar{u}, |\psi|^2) = 0$ . Thus, it follows from Lemma 2.2 that

$$\begin{aligned} \sigma((u_1, \psi_1), (u_2, \psi_2)) &= \nu(\nabla u_1, \nabla u_2) + b(\bar{u}, u_1, u_2) \\ &+ \kappa(\nabla \psi_1, \nabla \psi_2)_{L^2} + \frac{1}{\epsilon} (\psi_1, \psi_2)_\Gamma + (\bar{u} \cdot \nabla \psi_1, \psi_2) \end{aligned}$$

for  $(u_i, \psi_i)$ ,  $i = 1, 2$ , defines a bounded, elliptic sesquilinear form on  $(V_0 \times H^1(\Omega))^2$  with

$$\sigma((u, \psi), (u, \psi)) = \nu |\nabla u|^2 + \kappa |\nabla \psi|^2 + \frac{1}{\epsilon} |\psi|_{\Gamma}^2,$$

where we used the fact that for  $\epsilon > 0$

$$|\psi|_{H^1}^2 = \kappa |\nabla \psi|^2 + \frac{1}{2\epsilon} |\psi|_{\Gamma}^2$$

defines the equivalent norm on  $H^1(\Omega)$ . Thus, it follows from Lax-Milgram theorem that given  $(\bar{u}, \bar{\theta}) \in V_0 \times H^1(\Omega)$

$$\sigma((u, \theta), (v, \psi)) = (\gamma \bar{\theta} e_d + f, v) + \frac{1}{\epsilon} (g - T_0, \psi)_{\Gamma}, \quad \theta \in H^1(\Omega)$$

has a unique solution  $(u, \theta)$  in  $V_0 \times H^1(\Omega)$ . Moreover,

$$|\theta|_{H^1}^2 = |\nabla \theta|_{L^2}^2 + \frac{1}{2\epsilon} |\theta|_{L^2(\Gamma)}^2 \leq \frac{1}{2\epsilon} |g - T_0|_{L^2(\Gamma)}^2$$

and

$$|u|_{V_0} \leq \frac{M_3}{\nu} (\gamma |\bar{\theta}|_{L^2} + |f|_{L^2}),$$

where  $|\phi|_{H_0} \leq M_3 |\phi|_{V_0}$ ,  $\phi \in V_0$ . Define a solution map  $\Phi$  on  $V_0 \times H^1(\Omega)$  by  $\Phi(\bar{u}, \bar{\theta}) = (u, \theta)$ . Let  $C$  be the closed convex subspace of  $V_0 \times H^1(\Omega)$ , defined by

$$C = \{(\phi, \psi) : |\phi|_{V_0} \leq \frac{M_3}{\nu} \left( \frac{\gamma M_4}{\sqrt{2\epsilon}} |g - T_0|_{L^2(\Gamma)} + |f|_{L^2} \right) \text{ and } |\psi|_{H^1} \leq \frac{1}{\sqrt{2\epsilon}} |g - T_0|_{L^2(\Gamma)}\},$$

where  $|\theta|_{L^2} \leq M_4 |\theta|_{H^1}$ ,  $\theta \in H^1(\Omega)$ . It follows from the above estimates that  $\Phi$  maps from  $C$  into  $C$ . Moreover, it is shown in [IST2] that  $\Phi$  is continuous and compact. Hence by Schauder fixed point theorem  $\Phi$  has a fixed point in  $V_0 \times H^1(\Omega)$  that defines a solution to (2.5a) and (2.7).

Moreover, we assume that  $\bar{T}_1 \leq g \leq \bar{T}_2$  a.e. in  $\Gamma$ . Let  $\psi = \sup(\theta_\epsilon, \bar{T}_2 - T_0)$ . Then  $\psi \in H^1(\Omega)$  [Tr] and  $(u_\epsilon \cdot \nabla \theta_\epsilon, \psi) = -\frac{1}{2}(\nabla \cdot u_\epsilon, |\psi|^2) = 0$ . From (2.7)

$$\kappa |\nabla \psi|^2 + \frac{1}{\epsilon} (\theta_\epsilon - (g - T_0), \psi)_{\Gamma},$$

where  $(\theta_\epsilon - (g - T_0)) \psi \geq 0$  a.e. on  $\Gamma$ . Hence  $|\nabla \psi|_{L^2}^2 = 0$  and thus  $\theta_\epsilon \leq T_2 - T_0$ . Similarly, we can prove that  $\theta_\epsilon \geq \bar{T}_1 - T_0$  by choosing  $\psi = \inf(\bar{T}_1 - T_0, \theta_\epsilon)$ .  $\square$

**Lemma 2.4.** Suppose  $\bar{T}_1 \leq g \leq \bar{T}_2$  a.e. in  $\Gamma$ . Let  $(u_\epsilon, \theta_\epsilon) \in V_0 \times H^1(\Omega)$  be the solution to (2.5a) and (2.7). Then there exists a sequence  $\{\epsilon\}$  of positive numbers such that  $u_\epsilon$  converges weakly to  $u$  in  $V_0$  and  $\theta_\epsilon$  converges weakly  $*$  to  $\theta$  in  $L^\infty(\Omega)$ , where  $(u, \theta) \in V_0 \times L^\infty(\Omega)$  satisfies (2.5) with  $\bar{T}_1 - T_0 \leq \theta \leq \bar{T}_2 - T_0$  a.e. in  $\Omega$ .

**Proof:** Setting  $\psi = \theta_\epsilon$  in (2.7), we obtain

$$(2.8) \quad \epsilon \kappa |\nabla \theta_\epsilon|_{L^2}^2 + \frac{1}{2} |\theta_\epsilon|_{\Gamma}^2 \leq \frac{1}{2} |g - T_0|_{\Gamma}^2$$

Hence  $|\theta_\epsilon|_1^2$  is uniformly bounded in  $\epsilon > 0$ . Note that from (2.5a)

$$|u_\epsilon|_{V_0} \leq \frac{M_0}{\nu} (\gamma |\theta_\epsilon|_{L^2} + |f|_{L^2}).$$

It follows from Lemma 2.3 that  $\bar{T}_1 - T_0 \leq \theta_\epsilon \leq \bar{T}_2 - T_0$ . Hence  $|\theta_\epsilon|_{L^2} \leq \sqrt{\text{meas}(\Omega)} \times \max(|\bar{T}_1 - T_0|, |\bar{T}_2 - T_0|)$  and  $|u_\epsilon|_{V_0}$  is uniformly bounded in  $\epsilon > 0$ . Thus there exist  $(u, \theta, \hat{g}) \in V_0 \times H^1(\Omega) \times L^2(\Gamma)$  and a sequence  $\{\epsilon\}$  of positive numbers such that the trace of  $\theta_\epsilon + T_0$  on  $\Gamma$  converges weakly to  $\hat{g}$  in  $L^2(\Gamma)$ ,  $\theta_\epsilon$  converges to  $\theta$  weakly  $*$  in  $L^\infty(\Omega)$  and  $u_\epsilon$  converges to  $u$  weakly in  $V_0$ , as  $\epsilon \rightarrow 0^+$ . From (2.7), we have for  $\psi \in H^1(\Omega)$

$$(2.9) \quad (\theta_\epsilon - (g - T_0), \psi)_\Gamma + \epsilon (\kappa (\nabla \theta_\epsilon, \nabla \psi) + (u_\epsilon \cdot \nabla \theta_\epsilon, \psi)) = 0,$$

where it follows from (2.8) that

$$\epsilon |\nabla \theta_\epsilon|_{L^2}^2 \quad \text{is uniformly bounded in } \epsilon > 0$$

and

$$|(u_\epsilon \cdot \nabla \theta_\epsilon, \psi)| \leq M |u_\epsilon|_{V_0} |\nabla \theta_\epsilon|_{L^2} |\psi|_{H^1}$$

for some  $M > 0$ . Thus, from (2.9)

$$|((\theta_\epsilon + T_0) - g, \psi)_\Gamma| \leq \tilde{M} \sqrt{\epsilon} |\psi|_{H^1}$$

for some  $\tilde{M} > 0$  and all  $\psi \in H^1(\Omega)$ , which implies that  $\hat{g} = g$  and  $\theta_\epsilon \rightarrow g - T_0$  as  $\epsilon \rightarrow 0^+$  in  $H^{-1/2}(\Gamma)$ . Setting  $\psi = (-\Delta_0)^{-1} \phi$ ,  $\phi \in H_1 = L^2(\Omega)$  in (2.7) and by Green's formula we have

$$(2.10) \quad \kappa (\theta_\epsilon, \phi)_{L^2(\Omega)} - (\theta_\epsilon, u_\epsilon \cdot \nabla \psi)_{L^2(\Omega)} + \kappa (\theta_\epsilon, n \cdot \nabla \psi)_\Gamma = 0.$$

Note that  $\nabla \psi \in H^1(\Omega)$ ,  $n \cdot \nabla \psi \in H^{1/2}(\Gamma)$ , and

$$|(\theta_\epsilon, u \cdot \nabla \psi)_{L^2}| \leq |\theta|_{L^2} |u|_{L^4} |\nabla \psi|_{L^4}$$

Since  $H^1(\Omega)$  is compactly embedded into  $L^4(\Omega)$  it follows from (2.10) that  $(u, \theta) \in V_0 \times L^\infty(\Omega)$  satisfies (2.5b). Since  $(u_\epsilon, \theta_\epsilon)$  converges weakly to  $(u, \theta)$  in  $V_0 \times L^2(\Omega)$  and  $H^1(\Omega)$  is compactly embedded into  $L^4(\Omega)$ , it follows from Lemma 2.2 that  $(u, \theta) \in V_0 \times L^\infty(\Omega)$  satisfies (2.5a).  $\square$

**2.2. Necessary Optimality Condition.** Let us denote by  $S(g)$ , the solution set of (2.2) for  $g \in L^2(\Gamma)$ . Then, we have the existence of minimizer for (2.1)–(2.2).

**Theorem 2.5.** Consider the minimization problem (2.1)–(2.2) which is equivalently written as

$$\text{minimize } J(u, T - T_0, g) = \varphi(u, T - T_0) + \frac{\beta}{2} |g - T_0|_{L^2(\Gamma)}^2$$

$$\text{over } (u, T - T_0) \in S(g) \text{ and } g \in \mathcal{C},$$

where  $\mathcal{C}$  is a closed convex set in  $L^2(\Gamma)$ . Assume that  $\varphi(u, \theta)$  is coercive in  $\theta \in H_1 = L^2(\Omega)$  and satisfies

$$\varphi(z) : z = (u, T - T_0) \in V_0 \times L^2(\Omega) \rightarrow \mathbb{R}^+ \text{ is convex and lower semicontinuous} \\ \text{and } 0 \leq \varphi(z) \leq b_1 |z|_{V_0 \times L^2(\Omega)}^2 + b_2 \text{ for } b_1, b_2 \in \mathbb{R}^+.$$

Then Problem (2.1)–(2.2) has a solution.

**Proof:** It follows from Lemma 2.4 that  $S(g)$  is nonempty for  $g \in L^\infty(\Gamma)$ . Let  $(u_k, \theta_k) \in S(g_k)$ ,  $g_k \in \mathcal{C}$  be a minimizing sequence. Since  $\beta > 0$   $|g_k - T_0|_{L^2(\Gamma)}$  is uniformly bounded in  $k$ . From the assumption on  $\varphi$  we have  $|\theta_k|_{L^2}$  is uniformly bounded. Setting  $v = u_k$  in (2.5a), we obtain from Lemma 2.2

$$\nu |u_k|_{V_0}^2 \leq (\gamma |\theta_k|_{L^2} + |f|_{L^2}) |u_k|_{H_0}.$$

Hence,  $|u_k|_{V_0}$  is uniformly bounded and thus there exists a subsequence of  $\{k\}$ , which will be denoted by the same index, such that  $(u_k, \theta_k, g_k)$  converges weakly to  $(u, \theta, g) \in V_0 \times H_1 \times \mathcal{C}$  since  $V_0 \times H_1 \times L^2(\Gamma)$  is a Hilbert space and  $\mathcal{C}$  is closed and convex. Since  $H^1(\Omega)$  is compactly embedded into  $L^1(\Omega)$ , it follows from Lemma 2.2 that  $b(u_k, u_k, v) \rightarrow b(u, u, v)$  for  $v \in V_0$ . Note that

$$|(\theta, u \cdot \nabla \psi)| \leq |\theta|_{L^2} |u|_{L^4} |\nabla \psi|_{L^4} \leq M |\theta|_{L^2} |u|_{L^4} |\phi|_{L^2}$$

for  $\psi = (-\Delta_0)^{-1} \phi$ . Thus,  $(\theta_k, u_k \cdot \nabla \psi) \rightarrow (\theta, u \cdot \nabla \psi)$  for  $\psi = (-\Delta_0)^{-1} \phi$  and  $(u, \theta) \in S(g)$ . Now, since  $\varphi$  is convex and lower semicontinuous it follows from [ET] that  $(u, \theta, g)$  minimizes (2.1).  $\square$

Recall that  $W$  is the completion of  $H_1 = L^2(\Omega)$  with respect to norm  $|(-\Delta_0)^{-1} \phi|_{H_1}$ . Problem (2.1)–(2.2) is equivalently written as a constrained minimization on  $x = (u, T - T_0, g) \in X = V_0 \times H_1 \times L^2(\Gamma)$  with

$$\text{minimize } J(x) = \varphi(u, T - T_0) + \frac{\beta}{2} |g - T_0|_{L^2(\Gamma)}^2 \text{ over } x \in X$$

$$\text{subject to } c(x) = 0 \text{ and } g \in \mathcal{C}$$

where the equality constraint  $c : X \rightarrow Y = V_0^* \times W$  is defined by

$$(2.11) \quad \langle c(x), (v, \phi) \rangle = \nu (\nabla u, \nabla v) + b(u, u, v) - \gamma (\theta, v) \\ + \kappa (\theta, \phi) - (\theta, u \cdot \nabla \psi) + \kappa (g - T_0, u \cdot \nabla \psi)_\Gamma,$$

for  $(v, \phi) \in V_0 \times H_1$  with  $\psi = (-\Delta_0)^{-1} \phi$ . Assume that  $x^* = (u^*, \theta^*, g^*)$  denotes the optimal solution of (2.1)–(2.2). Then we have

**Theorem 2.6.** Assume that  $x^*$  is a regular point in the sense [MZ] that

$$(2.12) \quad 0 \in \text{int} \{c'(x^*)(v, \phi, h) : (v, \phi) \in V_0 \times H_1 \text{ and } h \in \mathcal{C} - g^*\} \text{ in } (V_0)^* \times W.$$

Then there exists a Lagrange multiplier  $(\lambda^*, \mu^*) \in V_0 \times (H^2(\Omega) \cap H_0^1(\Omega))$  such that

$$\begin{aligned} & \nu(\nabla \lambda^*, \nabla v) + b(v, u^*, \lambda^*) + b(u^*, v, \lambda^*) \\ (2.13) \quad & -(\theta^* \nabla \mu^*, v) + \langle \varphi_u(u^*, \theta^*), v \rangle = 0 \\ & \kappa(\nabla \mu^*, \nabla \psi) - \langle u^* \cdot \nabla \mu^*, \psi \rangle - \gamma(\lambda^*, \psi e_d) + \langle \varphi_\theta(u^*, \theta^*), \psi \rangle = 0 \end{aligned}$$

for  $v \in V_0$  and  $\psi \in H_0^1(\Omega)$  and

$$(2.14) \quad (\beta(g^* - T_0) + \kappa n \cdot \nabla \mu^*, h - g^*)_{\Gamma} \geq 0 \quad \text{for all } h \in \mathcal{C}.$$

**Proof:** It follows from Lemma 2.2 and (2.11) that  $c$  is Fréchet differentiable and the  $F$ -derivative  $c'(x^*)(v, h)$  is given by

$$\begin{aligned} \langle c'(x^*)(v, \phi, h), \psi \rangle &= \nu(\nabla v, \nabla \psi_1) + b(v, u^*, \psi_1) + b(u^*, v, \psi_1) - \gamma(\phi e_d, \psi_1) \\ &\quad + \kappa(\phi, \psi_2) - (\theta^*, v \cdot \nabla \psi) - \langle \phi, u^* \cdot \nabla \psi \rangle + \kappa(h, n \cdot \psi)_{\Gamma} \end{aligned}$$

for  $(v, \phi, h) \in X$  and  $\psi = (\psi_1, \psi_2) \in V_0 \times H_1$ , where  $\psi = (-\Delta_0)^{-1} \psi_2$ . Since  $x^*$  is regular, it then follows from [MZ] that there exists a Lagrange multiplier  $\lambda = (\lambda^*, \eta^*) \in Y^* = V_0 \times H_1$  such that

$$(2.15) \quad \langle \varphi'(u^*, \theta^*), (v, \phi) \rangle + \beta(g^* - T_0, h - g^*)_{\Gamma} + \langle c'(x^*)(v, \phi, h - g^*), \lambda \rangle \geq 0$$

for all  $(v, \phi) \in V_0 \times H_1$  and  $h \in \mathcal{C}$ . Setting  $(v, \phi) = 0$ , we obtain (2.14) if we define

$$\mu^* = (-\Delta_0)^{-1} \eta^*.$$

Next, setting  $h = g^*$  in (2.15), we have

$$\langle \varphi'(u^*, \theta^*), (v, \phi) \rangle + \langle c'(x^*)(v, \phi, 0), \lambda \rangle = 0$$

for all  $(v, \phi) \in V_0 \times H_1$ . Thus we obtain

$$\nu(\nabla \lambda^*, \nabla v) + b(v, u^*, \lambda^*) + b(u^*, v, \lambda^*) - (\theta^* \nabla \mu^*, v) + \langle \varphi_u(u^*, \theta^*), v \rangle = 0$$

for all  $v \in V_0$  and

$$\kappa(-\Delta_0 \mu^*, \phi)_{L^2} - \langle u^* \cdot \nabla \mu^*, \phi \rangle - \gamma(\lambda^*, \phi e_d) + \langle \varphi_\theta(u^*, \theta^*), \phi \rangle = 0,$$

for all  $\phi \in H_1 = L^2(\Omega)$  which implies (2.13).  $\square$

Note that if the linear operator  $E \in \mathcal{L}(V_0 \times H_1, V_0^* \times W)$ , defined by  $E(v, \phi) = c'(x^*)(v, \phi, 0)$ , is surjective, then the regular point condition (2.12) is satisfied. Moreover, we have the following lemma.

**Lemma 2.7.** *If  $g^* \in \text{int}(\mathcal{C})$  then the regular point condition (2.12) is equivalent to that equation for  $(\lambda, \mu) \in V_0 \times (H^2(\Omega) \cap H_0^1(\Omega))$*

$$\begin{aligned} & \nu(\nabla \lambda, \nabla v) + b(v, u^*, \lambda) + b(u^*, v, \lambda) - (\theta^* \nabla \mu, v) = 0 \quad \text{for } v \in V_0 \\ (2.16) \quad & \kappa(\nabla \mu, \nabla \psi) - (u^* \cdot \nabla \mu, \psi) - \gamma(\lambda, \psi c_d) = 0, \quad \text{for } \psi \in H_0^1(\Omega) \\ & \text{and } n \cdot \nabla \mu = 0 \quad \text{on } \Gamma, \end{aligned}$$

implies  $(\lambda, \mu) = 0$ .

**Proof:** If  $g^* \in \text{int}(\mathcal{C})$  then (2.12) is equivalent to that  $G = c'(x^*)$  is surjective. Define a linear map  $C \in \mathcal{L}(X, V_0 \times H_1)$  by  $C(v, \phi, h) = (U, \Theta)$  where  $(U, \Theta) \in V_0 \times H_1$  is the unique solution to

$$\begin{aligned} & \nu(\nabla U, \nabla \psi_1) + b(v, u^*, \psi_1) + b(u^*, v, \psi_1) - \gamma(\phi c_d, \psi_1) = 0 \\ & \kappa(\Theta, \psi_2) - (\phi, u^* \cdot \nabla \psi) - (\theta^*, v \cdot \nabla \psi) + \kappa(h, n \cdot \nabla \psi)_\Gamma = 0 \end{aligned}$$

for  $(\psi_1, \psi_2) \in V_0 \times H_1$  and  $\psi = (-\Delta_0)^{-1} \psi_2$ . Since  $H^1(\Omega)$  is compactly embedded into  $L^4(\Omega)$ , it follows from Lemma 2.2 that  $C$  is compact. It thus follows from Banach closed range and Riesz-Schauder theorems that  $c'(x^*)$  is surjective if and only if  $\ker(C^*) = \{0\}$  [DI], which is equivalent to (2.16).  $\square$

### 3. Finite-time horizon problem

In this section we discuss the finite-time horizon problem (1.1)-(1.2). First, we formulate the weak form of equation (1.1)-(1.2). By Green's formula we have from (1.2)

$$\begin{aligned} \left\langle \frac{d}{dt} \theta(t), \psi \right\rangle &= (-u(t) \cdot \nabla \theta(t) + \kappa \Delta \theta(t), \psi) \\ &= -\kappa(\theta(t), \phi) - \kappa(g(t) - T_0, n \cdot \nabla \psi)_\Gamma + (\theta(t), u(t) \cdot \nabla \psi) \end{aligned}$$

for  $\psi = (-\Delta_0)^{-1} \phi$ ,  $\phi \in H_1 = L^2(\Omega)$ , where we assumed that  $\theta(t)$  is sufficiently smooth. We define the weak solution of (1.1)-(1.2) by

**Definition 3.1.** The pair  $(u, \theta) \in L^2(0, T; V_0 \times H_1) \cap W^{1,s}(0, T; V_0^* \times W)$  with  $s > 1$  is a weak solution to (1.1)-(1.2) if

$$(3.1a) \quad \left\langle \frac{d}{dt} u(t), v \right\rangle_{V_0^* \times V_0} + \nu(\nabla u(t), \nabla v) + b(u(t), u(t), v) - \gamma(\theta(t) c_d, v) = 0$$

for all  $v \in V_0$ ,

$$(3.1b) \quad \left\langle \frac{d}{dt} \theta(t), \phi \right\rangle_{W \times H_1} - \kappa(\theta(t), \phi) - (\theta(t), u \cdot \nabla \psi) + \kappa(g(t) - T_0, n \cdot \nabla \psi)_\Gamma = 0$$

for all  $\phi \in H_1$ , where  $\psi = (-\Delta_0)^{-1} \phi$ .

Recall that  $H_1 = L^2(\Omega)$ ,  $V_1^* = H^{-1}(\Omega)$  and  $V_0 \times H_1 \subset H_0 \times V_1^* = (H_0 \times V_1^*)^* \subset V_0^* \times W$  defines the Gelfand triple where  $H_0 \times V_1^*$  is the pivoting space and  $V_1^*$  is equipped with inner product (2.3).

Again, we construct a solution to (3.1) by the limiting procedure for solutions to (1.5) as  $\epsilon \rightarrow 0^+$ . To this end we consider equation (1.5). Let  $V = V_0 \times H^1(\Omega)$  and  $H = H_0 \times L^2(\Omega)$ . For  $\epsilon > 0$  define a sesquilinear form  $\hat{a}$  on  $V \times V$  by

$$(3.2) \quad \hat{a}(\phi, \psi) = \nu(\nabla \phi_1, \nabla \psi_1) - \gamma(\phi_2 e_d, \psi_1) + \kappa(\nabla \phi_2, \nabla \psi_2) + \frac{1}{\epsilon}(\phi_2, \psi_2)_\Gamma$$

for  $\phi = (\phi_1, \phi_2)$ ,  $\psi = (\psi_1, \psi_2) \in V$ . Then,  $\hat{a}$  satisfies

$$(3.3) \quad \begin{aligned} |\hat{a}(\phi, \psi)| &\leq M|\phi|_V|\psi|_V \text{ for } \phi, \psi \in V \text{ and} \\ \hat{a}(\phi, \phi) &\geq \omega|\phi|_V^2 - \gamma|\phi|_H^2 \text{ for } \phi \in V, \end{aligned}$$

for  $\omega > 0$  and  $M \geq 1$ , where we used the fact that

$$(3.4) \quad \kappa|\nabla \theta|_{L^2(\Omega)}^2 + \frac{1}{2\epsilon}|\theta|_{L^2(\Gamma)}^2 \geq \omega|\theta|_{H^1(\Omega)}^2$$

for  $\epsilon \leq 1$ . Define the tri-linear form  $\tilde{b}$  on  $V^3$  by

$$\tilde{b}((u, \theta_1), (v, \theta_2), (w, \theta_3)) = b(u, v, w) + (u \cdot \nabla \theta_2, \theta_3)$$

Since

$$(3.5) \quad (u \cdot \nabla \theta_1, \nabla \theta_2) + (u \cdot \nabla \theta_1, \nabla \theta_2) = -(\nabla \cdot u, \theta_1 \theta_2) = 0$$

for  $u \in V_0$  and  $\theta_1, \theta_2 \in H^1(\Omega)$ , it follows that  $\tilde{b}$  satisfies Lemma 3.2 in which  $V_0, H_0$  are replaced by  $V, H$ , respectively. The weak form of equation (1.5) is then given by

$$(3.6a) \quad \left\langle \frac{d}{dt} u(t), \phi \right\rangle_{V_0^* \times V_0} + \nu(\nabla u(t), \nabla \phi) + b(u(t), u(t), \phi) = (\gamma \theta(t) e_d + f, \phi),$$

for  $\phi \in V_0$  and

$$(3.6b) \quad \left\langle \frac{d}{dt} \theta(t), \psi \right\rangle_{(H^1)^* \times H^1} + \kappa(\nabla \theta(t), \nabla \psi) + \frac{1}{\epsilon}(\theta(t) - (g(t) - T_0), \psi)_\Gamma = 0$$

for  $\psi \in H^1(\Omega)$ , where  $\theta(t) = T(t, \cdot) - T_0$ , or equivalently

$$\left\langle \frac{d}{dt} z(t), \psi \right\rangle_{V \times V} + \hat{a}(z(t), \psi) + \tilde{b}(z(t), z(t), \psi) = (f, \psi_1) + \frac{1}{\epsilon}(g(t) - T_0, \psi_2)_\Gamma,$$

for  $\psi = (\psi_1, \psi_2) \in V$ , where  $z(t) = (u(t, \cdot), \theta(t))$ . Note that

$$|(g, \psi)_\Gamma| \leq M|g|_{L^2(\Gamma)}|\psi|_{H^1(\Omega)}$$

for  $g \in L^2(\Gamma)$  and  $\psi \in H^1(\Omega)$ . Hence, by using the standard arguments in [Tel], [CF] we have

**Theorem 3.2.** *If  $d = 2$  then for any  $z(0) \in H$  and  $g \in U = L^2(0, T; L^2(\Gamma))$ , equation (1.3)–(1.4) has a unique solution  $z(t) = (u(t, \cdot), \theta(t) = T(t, \cdot) - T_0) \in$*



$C(0, T; H) \cap L^2(0, T; V) \cap H^1(0, T; V^*)$  satisfying (3.6) a.e. in  $(0, T)$  for all  $T$ . If  $d = 3$  then equation (1.3)–(1.4) has a weak solution  $z(t) \in C_w(0, T; H) \cap L^2(0, T; V) \cap W^{1,4/3}(0, T; V^*)$  for any  $z(0) \in H$  and  $g \in L^2(0, T; L^2(\Gamma))$  with bound

$$(3.7) \quad \begin{aligned} & \frac{1}{2} |z(t)|_H^2 + \int_0^t \nu |u(s)|_{V_0}^2 + \kappa |\nabla \theta(s)|^2 + \frac{1}{\epsilon} |\theta(s)|_\Gamma^2 ds \\ & \leq \frac{1}{2} |z(0)|_H^2 + \int_0^t (\gamma \theta(s) c_d + f, u(s)) + \frac{1}{\epsilon} (\theta(s), (g(s) - T_0) - \theta(s))_\Gamma ds \end{aligned}$$

for  $t \geq 0$ . In (3.7) equality holds instead of the inequality when  $d = 2$ .

Let  $K$  be the closed convex set in  $L^2(0, T; L^2(\Gamma))$  defined by

$$(3.8) \quad K = \{g(t) \in L^2(0, T; L^2(\Gamma)) : g(t) \in \mathcal{C} \text{ a.e. in } (0, T)\}.$$

Then, the following corollary is proved in [IST2].

**Corollary 3.3.** Assume that  $g \in K$ , i.e.,  $T_1 - T_0 \leq g(t, \cdot) \leq T_2 - T_0$  a.e. in  $(0, T) \times \Omega$  and  $T_1 \leq T(0, \cdot) \leq T_2$  a.e. in  $\Omega$ . Then, there exists a solution to (3.6) such that  $T_1 - T_0 \leq \theta(t, \cdot) \leq T_2 - T_0$  a.e. in  $(0, T) \times \Omega$ . For such a solution  $\theta(t) = T(t) - T_0 \in C(0, T; L^2(\Omega)) \cap L^2(0, T; H^1(\Omega)) \cap H^1(0, T; H^1(\Omega)^*)$ .

For  $0 < \epsilon \leq 1$  let us denote by  $(u_\epsilon, \theta_\epsilon)$  the solution to (3.6) in the sense of Theorem 3.2 and Corollary 3.3. Next, we show that  $(u_\epsilon, \theta_\epsilon)$  converges to a solution (3.1) in the sense of Definition 3.1.

**Theorem 3.4.** Suppose  $g \in K$  and  $T_1 \leq T(0, \cdot) \leq T_2$  a.e. in  $\Omega$ . Then there exists a sequence  $\{\epsilon\}$  of positive numbers such that  $u_\epsilon$  converges to  $u$  weakly in  $L^2(0, T; V_0)$  and  $W^{1,s}(0, T; V_0^*)$  ( $s = 2$  if  $d = 2$  and  $s = 4/3$  if  $d = 3$ ) and strongly in  $L^2(0, T; H_0)$ , and  $\theta_\epsilon$  converges to  $\theta$  in weakly  $*$  in  $L^\infty((0, T) \times \Omega)$  and strongly in  $L^2(0, T; H^{-1}(\Omega))$  as  $\epsilon \rightarrow 0^+$ , where the pair  $(u, \theta)$  is a solution to (3.1) with  $T_1 - T_0 \leq \theta(t, \cdot) \leq T_2 - T_0$  a.e. in  $(0, T) \times \Omega$ .

**Proof:** The proof is given in several steps.

**Step 1.** We establish the uniform bound for  $(u_\epsilon, \theta_\epsilon)$  and the trace of  $\theta_\epsilon(t)$  on  $\Gamma$ . Since for  $\theta(t) \in L^2(0, T; H^1(\Omega)) \cap H^1(0, T; H^1(\Omega)^*)$

$$(3.9) \quad \frac{1}{2} \frac{d}{dt} |\theta(t)|_{L^2}^2 = \left\langle \frac{d}{dt} \theta(t), \theta(t) \right\rangle_{(H^1)^*, H^1} \quad \text{a.e. } t \in (0, T),$$

it follows from Corollary 3.3 and (3.6b) that

$$\frac{1}{2} \frac{d}{dt} |\theta_\epsilon(t)|_{L^2}^2 + \kappa |\nabla \theta_\epsilon(t)|^2 + \frac{1}{2\epsilon} |\theta_\epsilon(t)|_\Gamma^2 \leq \frac{1}{2\epsilon} |g(t) - T_0|_\Gamma^2$$

Integrating this with respect to  $t$ , we obtain

$$(3.10) \quad |\theta_\epsilon(t)|_{L^2}^2 + \int_0^t 2\kappa |\nabla \theta_\epsilon(s)|^2 + \frac{1}{\epsilon} |\theta_\epsilon(s)|_\Gamma^2 ds \leq |\theta(0)|_{L^2}^2 + \int_0^t \frac{1}{\epsilon} |g(s) - T_0|_\Gamma^2 ds$$

for  $t \in [0, T]$ . Hence  $\int_0^T |\theta_\epsilon(s)|_1^2 ds$  is uniformly bounded in  $\epsilon > 0$ .

From (3.7) and (3.9) we have

$$|u_\epsilon(t)|_{H_0}^2 + \int_0^t \nu |u_\epsilon(s)|_{V_0}^2 ds \leq \frac{M_3}{\nu} \int_0^t \gamma |\theta_\epsilon(s)|_{L^2}^2 + |f|_{L^2}^2 ds,$$

where  $|\phi|_{H_0} \leq M_3 |\phi|_{V_0}$ . It follows from Corollary 3.3 that

$$(3.11) \quad \|\theta_\epsilon(t)\|_{L^\infty(\Omega)} \leq \max(\bar{T}_1 - T_0, \bar{T}_2 - T_0).$$

Hence  $\int_0^T |\theta_\epsilon(s)|_{L^2(\Omega)}^2 ds$  is uniformly bounded and so is

$$\sup_{t \in [0, T]} |u_\epsilon(t)|_{H_0}^2 + \int_0^T \nu |u_\epsilon(t)|_{V_0}^2 dt$$

in  $\epsilon > 0$ . It follows from Lemma 2.2 that

$$|b(u(t), u(t), \phi)| \leq M_2 |u(t)|_{V_0} |u(t)|_{H_0} |\phi|_{V_0} \quad \text{for } d = 2$$

and

$$|b(u(t), u(t), \phi)| \leq M_2 |u(t)|_{V_0}^{3/2} |u(t)|_{H_0}^{1/2} |\phi|_{V_0} \quad \text{for } d = 3.$$

Hence, from (3.6a)

$$\left| \frac{d}{dt} u_\epsilon(t) \right|_{V_0} \leq \nu |u_\epsilon(t)|_{V_0} + M_2 |u_\epsilon(t)|_{V_0}^{3/2} |u_\epsilon(t)|_{H_0}^{1/2} + M_3 (\gamma |\theta_\epsilon(t)|_{L^2} + |f|_{L^2}),$$

and  $\int_0^T \left| \frac{d}{dt} u_\epsilon(t) \right|_{V_0}^{4/3} dt$  is uniformly bounded for  $d = 3$ . Similarly, for  $d = 2$  we have  $\int_0^T \left| \frac{d}{dt} u_\epsilon(t) \right|_{V_0}^2 dt$ .

Setting  $\psi = (-\Delta_0)^{-1} \phi$ ,  $\phi \in H_1 = L^2(\Omega)$  in (3.6b) and by Green's formula we obtain

$$(3.12) \quad \left\langle \frac{d}{dt} \theta_\epsilon(t), \phi \right\rangle_{W^*, H_1} + \kappa (\theta_\epsilon(t), \phi)_{L^2} - (\theta_\epsilon(t), u_\epsilon(t) \cdot \nabla \psi)_{L^2} \\ + \kappa (\theta_\epsilon(t), u_\epsilon \cdot \nabla \psi)_\Gamma = 0.$$

Since

$$|(\theta_\epsilon(t), u_\epsilon(t) \cdot \nabla \psi)_{L^2}| \leq M \|\theta_\epsilon(t)\|_{L^2} \|u_\epsilon(t)\|_{V_0} \|\phi\|_{L^2}$$

for some  $M > 0$  it follows from (3.12) that  $\int_0^T \left| \frac{d}{dt} \theta_\epsilon(t) \right|_{W^*}^2 dt$  is uniformly bounded.

**Step 2.** We show that  $(u_\epsilon(t), \theta_\epsilon(t))$  and the trace of  $\theta_\epsilon(t) + T_0$  on  $\Gamma$  have the appropriate limit. It follows from Aubin's lemma [CF] that there exist  $g(t) \in$

$L^2(0, T; L^2(\Gamma))$ ,  $u(t) \in L^2(0, T; V_0) \times W^{1,s}(0, T; V_0^*)$ ,  $\theta(t) \in L^\infty((0, T) \times \Omega) \times H^1(0, T; W)$  and a subsequence  $\{\epsilon\}$  of positive numbers such that

$(\theta_\epsilon(t) + T_0)|_\Gamma$  converges weakly to  $\hat{g}(t)$  in  $L^2(0, T; L^2(\Gamma))$ ,

$u_\epsilon(t) \rightharpoonup u(t)$  weakly in  $L^2(0, T; V_0)$  and strongly in  $L^2(0, T; H_0)$

$\theta_\epsilon(t) \rightharpoonup \theta(t)$  weakly  $*$  in  $L^\infty((0, T) \times \Omega)$  and strongly in  $L^2(0, T; H^{-1}(\Omega))$

From (3.6b), for  $\psi(\cdot) \in L^2(0, T; H^1(\Omega))$

$$\begin{aligned} \epsilon \left\langle \frac{d}{dt} \theta_\epsilon(t), \psi(t) \right\rangle + (\theta_\epsilon(t) - (g(t) + T_0), \psi(t))_\Gamma \\ + \epsilon \kappa (\nabla \theta_\epsilon(t), \nabla \psi(t)) + \epsilon (u_\epsilon \cdot \nabla \theta_\epsilon(t), \psi(t)) = 0 \end{aligned}$$

where from (3.10)

$$\epsilon \int_0^T \left| \frac{d}{dt} \theta_\epsilon(t) \right|_{(H^1)^*}^2 + \kappa \|\nabla \theta_\epsilon(t)\|_{L^2}^2 dt \quad \text{is uniformly bounded in } \epsilon > 0$$

and from (3.5) and (3.11)

$$\left| \int_0^T (u_\epsilon(t) \cdot \nabla \theta_\epsilon(t), \psi(t)) dt \right| \leq \mathcal{A} \|u_\epsilon\|_{L^2(0,T;H_0)} \|\psi\|_{L^2(0,T;H^1)},$$

with  $\mathcal{A} = \max(|T_1 - T_0|, |T_2 - T_0|)$ . Thus,

$$(3.13) \quad \left| \int_0^T (\theta_\epsilon(t) + T_0 - g(t), \psi(t))_\Gamma dt \right| \leq \tilde{M} \sqrt{\epsilon} \|\psi(t)\|_{L^2(0,T;H^1)}$$

for some  $\tilde{M} > 0$  and all  $\psi \in L^2(0, T; H^1(\Omega))$ , which implies that  $\hat{g}(t) = g(t)$  and  $\theta_\epsilon(t) + T_0 \rightharpoonup g(t)$  as  $\epsilon \rightarrow 0^+$  in  $L^2(0, T; H^{-1/2}(\Gamma))$ .

**Step 3.** We show that  $(u(t), \theta(t))$  is a weak solution of (1.2) in the sense of Definition 3.1. From (3.12)

$$\begin{aligned} ((-\Delta_0)^{-1}(\theta_\epsilon(t) - \theta(0)), \phi)_{L^2} \\ (3.14) \quad + \int_0^t \kappa (\theta_\epsilon(t), \phi)_{L^2} - (\theta_\epsilon(t), u_\epsilon(t) \cdot \nabla \phi)_{L^2} + \kappa (\theta_\epsilon(t), u_\epsilon \cdot \nabla \phi)_\Gamma dt = 0 \end{aligned}$$

for  $\phi = (-\Delta_0)^{-1} \phi$ ,  $\phi \in L^2(\Omega)$ . Since every  $L^2$ -convergent sequence has an almost everywhere pointwise convergent subsequence, we can assume that

$$(u_\epsilon(t), (-\Delta_0)^{-1} \theta_\epsilon(t)) \rightharpoonup (u(t), (-\Delta_0)^{-1} \theta(t)) \quad \text{in } L^2(\Omega)^d \times L^2(\Omega)$$

for  $t \in (0, T] \setminus E$ , where  $E$  has zero Lebesgue measure. From (3.6a)

$$(3.15) \quad \begin{aligned} & (u_\epsilon(t) - u(0), v) \\ & + \int_0^t \nu (\nabla u_\epsilon(s), \nabla v) + b(u_\epsilon(s), u_\epsilon(s), v) - (\gamma \theta_\epsilon(s) e_d + f, v) ds = 0 \end{aligned}$$

for  $v \in V_0$ . Note that

$$\begin{aligned} & \int_0^t |b(u_\epsilon(s) - u(s), u_\epsilon(s), v)| ds \\ & \leq M \int_0^t |u_\epsilon(s) - u(s)|_{H_0}^{1/2} |u_\epsilon(s) - u(s)|_{V_0}^{1/2} |u_\epsilon(s)|_{V_0} |v|_{V_0} ds \\ & \leq M |u_\epsilon(s) - u(s)|_{L^2(0, T; H_0)}^{1/2} |u_\epsilon(s) - u(s)|_{L^2(0, T; V_0)}^{1/2} |u_\epsilon(s)|_{L^2(0, T; V_0)} |v|_{V_0} \end{aligned}$$

where we used the fact that  $|\phi|_{L^6} \leq c|\phi|_{H^1}$  and  $|\phi|_{L^3} \leq c|\phi|_{L^2}^{1/2}|\phi|_{H^1}^{1/2}$  for  $\phi \in H^1(\Omega)$  and the Hölder inequality. Similarly, we have

$$\begin{aligned} & \int_0^t |(\theta_\epsilon(s), (u_\epsilon(s) - u(s)) \cdot \nabla \psi)| ds \\ & \leq M |u_\epsilon(s) - u(s)|_{L^2(0, T; H_0)}^{1/2} |u_\epsilon(s) - u(s)|_{L^2(0, T; V_0)}^{1/2} |\theta_\epsilon(s)|_{L^2(0, T; L^2)} |v|_{H^2}. \end{aligned}$$

Hence, we can pass the limit of  $\epsilon \rightarrow 0^+$  in (3.14) and (3.15) to obtain

$$(3.16) \quad \begin{aligned} & (u(t) - u(0), v) \\ & + \int_0^t \nu (\nabla u(s), \nabla v) + b(u(s), u(s), v) - (\gamma \theta(s) e_d + f, v) ds = 0 \\ & (\theta(t) - \theta(0), \psi) \\ & + \int_0^t \kappa(\theta(t), \phi)_{L^2} - (\theta(t), u(t) \cdot \nabla \psi) + \kappa(\theta(t), u \cdot \nabla \psi)_T dt = 0 \end{aligned}$$

for  $v \in V_0$ ,  $\psi = (-\Delta_0)^{-1}\phi$ ,  $\phi \in L^2(\Omega)$  and  $t \in (0, T] \setminus E$ . Since the integrands appearing in (3.16) are integrable, (3.16) holds for all  $t \in (0, T]$ . Therefore  $(u, \theta)$  is a weak solution to (1.2).  $\square$

**Corollary 3.5.** Suppose  $u(0) \in V_0$  and  $d = 2$ . Then the solution to (1.1)–(1.2) in the sense of Definition 3.1 is unique and  $(u_\epsilon, \theta_\epsilon) \rightharpoonup (u, \theta)$  in  $L^2(0, T; V_0 \times L^2(\Omega)) \cap H^1(0, T; V_0^* \times W)$ .

**Proof:** Since  $\theta(t) \in L^2(0, T; L^2(\Omega))$  it follows from [Te1], [WvW] that we have  $u(t) \in C(0, T; V_0) \cap L^2(0, T; H^2(\Omega) \cap V_0)$ . Note that

$$(3.17) \quad \begin{aligned} & |(\theta_1, u_1 \cdot \nabla(-\Delta_0)^{-1}\phi) - (\theta_2, u_2 \cdot \nabla(-\Delta_0)^{-1}\phi)| \\ & \leq |\theta_1 - \theta_2|_{L^2} |u_2|_{H^2} |\phi|_{H^{-1}} + |\theta_1|_{L^2} |u_2 \\ & \quad - u_1|_{V_0}^{1/2} |u_2 - u_1|_{H_0}^{1/2} |\phi|_{H^{-1}}^{1/2} |\phi|_{L^2}^{1/2} \end{aligned}$$

for  $\theta_i \in L^2(\Omega)$ ,  $u_i \in H^2(\Omega) \cap V_0$ ,  $i = 1, 2$  and  $\phi \in L^2(\Omega)$ , where we used the fact that  $|\psi|_{L^1} \leq c |\psi|_{L^2}^{1/2} |\psi|_{H^1}^{1/2}$  for  $\psi \in H^1(\Omega)$ . Suppose  $(u_i, \theta_i)$ ,  $i = 1, 2$  is two solutions to (3.1). Since

$$|(\theta(t), u(t) \cdot \nabla(-\Delta_0)^{-1}\phi)| \leq M |\theta(t)|_{L^2} |u(t)|_{V_0} |\phi|_{L^2},$$

we have  $\theta_i(t) \in H^1(0, T; W)$ . Then, it follows from Lemma 2.2 and (3.17) that

$$\begin{aligned} & |u_1(t) - u_2(t)|_{H_0}^2 + |\theta_1(t) - \theta_2(t)|_{H^{-1}}^2 \\ & \leq M \int_0^t (1 + |u_1(s)|_{V_0}^2 + |\theta_1(s)|_{L^2}^2) |u_1(s) - u_2(s)|_{H_0}^2 + |u(s)|_{H^2}^2 |\theta_1 - \theta_2(s)|_{H^{-1}}^2 ds. \end{aligned}$$

Hence, by Gronwall's inequality we have  $(u_1(t), \theta_1(t)) = (u_2(t), \theta_2(t))$  for every  $t \in [0, T]$ .

Let  $v(t) = u_t(t) - u(t)$  and  $w(t) = \theta_t - \theta(t)$ . Since  $(v(t), w(t)) \in L^2(0, T; V_0 \times L^2(\Omega)) \cap H^1(0, T; V_0^* \times W)$ , it follows from (3.1) and (3.12) that

$$\begin{aligned} & \frac{1}{2} \frac{d}{dt} |v(t)|_{H_0}^2 + \nu |v(t)|_{V_0}^2 + b(v(t), u(t), v(t)) = \gamma(w(t) e_d, v(t)) \\ & \frac{1}{2} \frac{d}{dt} |w(t)|_{H^{-1}}^2 + \kappa |w(t)|_{L^2}^2 + (\theta_t(t), v(t) \cdot \nabla(-\Delta_0)^{-1}w(t)) \\ & + (w(t), u_t(t) \cdot \nabla(-\Delta_0)^{-1}w(t)) + \kappa (\theta_t + T_0 - g(t), n \cdot \nabla(-\Delta_0)^{-1}w(t))_{\Gamma} = 0. \end{aligned}$$

Note that

$$|(\theta_t(t) + T_0 - g(t), n \cdot \nabla(-\Delta_0)^{-1}w(t))_{\Gamma}| \leq M |(\theta_t(t) + T_0) - g(t)|_{H^{-1/2}(\Gamma)} |w(t)|_{L^2}$$

and that from (3.13)  $\theta_t + T_0 \rightarrow g(t)$  in  $L^2(0, T; H^{-1/2}(\Gamma))$ . Hence, from (3.17)

$$\begin{aligned} & |v(t)|_{H_0}^2 + |w(t)|_{H^{-1}}^2 + \int_0^t \nu |v(s)|_{V_0}^2 + \kappa |w(s)|_{L^2}^2 ds \\ & \leq M \int_0^t (1 + |u_1(s)|_{V_0}^2 + |\theta_1(s)|_{L^2}^2) |v(s)|_{H_0}^2 + |u(s)|_{H^2}^2 |w(s)|_{H^{-1}}^2 \\ & \quad + \kappa |(\theta_t(s) + T_0) - g(s)|_{H^{-1/2}(\Gamma)}^2 ds \end{aligned}$$

The desired convergence property follows from Gronwall's inequality.  $\square$

Let us denote by  $S(g)$ , the solution set of (1.2) for  $g \in L^2(\Gamma)$  in the sense of Definition 3.1. Then, we have

**Theorem 3.6.** *Consider the finite-time horizon problem (1.1)–(1.2) which is equivalently written as*

$$\begin{aligned} \text{minimize } J(u, T - T_0, g) &= \int_0^T \varphi(u(t), T(t) - T_0) + \frac{\beta}{2} \|g(t) - T_0\|_{L^2(\Gamma)}^2 dt \\ \text{over } (u(t), T(t) - T_0) &\in S(g) \text{ and } g(t) \in K. \end{aligned}$$

where  $K$  is a closed convex set in  $U = L^2(0, T; L^2(\Gamma))$ . Assume that  $\varphi(u, \theta)$  is coercive in  $\theta \in H_1 = L^2(\Omega)$  and satisfies

$$\begin{aligned} \varphi(z) : z = (u, T - T_0) \in V_0 \times L^2(\Omega) &\rightarrow R^+ \text{ is convex and lower semicontinuous} \\ \text{and } 0 \leq \varphi(z) &\leq b_1 \|z\|_{V_0 \times L^2}^2 + b_2 \text{ for } b_1, b_2 \in R^+. \end{aligned}$$

Then Problem (1.1)–(1.2) has a solution.

**Proof:** It follows from Lemma 3.4 that  $S(g)$  is nonempty for  $g \in L^\infty((0, T) \times \Gamma)$ . Let  $(u_k, g_k) \in S(g_k) \times K$  be a minimizing sequence. Since  $\beta > 0$   $\|g_k(t) - T_0\|_U$  is uniformly bounded in  $k$ . It follows from the assumption on  $\varphi$  that  $\|\theta_k(t)\|_{L^2(0, T; L^2(\Omega))}$  is uniformly bounded. From (3.1a) we have

$$\|u_k(t)\|_{H_0}^2 + \int_0^t \nu \|u_k(s)\|_{V_0}^2 ds \leq \frac{M_3^2}{\nu} \int_0^t \gamma^2 \|\theta_k(s)\|_{L^2}^2 + \|f\|_{L^2}^2 ds$$

for  $t \in [0, T]$ . Hence  $\sup_{t \in [0, T]} \|u_k(t)\|_{H_0}^2 + \int_0^T \|u_k(t)\|_{V_0}^2 dt$  is uniformly bounded. This implies that from Lemma 2.2  $(u_k, \theta_k) \in W^{1,1/3}(0, T; V_0^* \times W)$ . It thus follows from Aubin's lemma that there exists a subsequence of  $\{k\}$ , which will be denoted by the same index, such that  $(u_k, \theta_k, g_k)$  converges to  $(u, \theta, g)$  weakly in  $L^2(0, T; V_0) \times L^2(0, T; L^2(\Omega)) \times K$  and strongly in  $L^2(0, T; H_0 \times H^{-1})$  since  $K$  is closed and convex. Hence, using the same arguments as in the proof of Theorem 3.4, it can be shown that  $(u(t), \theta(t), g(t))$  satisfies (3.1). Now, since  $\varphi$  is convex and lower semicontinuous it follows from [ET] that  $(u, \theta, g)$  minimizes (1.1).  $\square$

We have the following necessary optimality condition.

**Theorem 3.7.** *Let  $d = 2$ ,  $K$  be given by (3.8) and  $u(0) \in V_0$ . If  $(u^*, \theta^*) \in S(g^*)$ ,  $g^* \in K$  minimizes (1.1), then there exists a Lagrange multiplier  $(\lambda(t), \mu(t)) \in L^2(0, T; V_0) \cap H^1(0, T; V_0^*) \times L^2(0, T; H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(0, T; L^2(\Omega))$ .*

such that

$$\begin{aligned}
 & \left\langle -\frac{d}{dt}\lambda(t), v \right\rangle + \nu (\nabla \lambda(t), \nabla v) + b(v, u^*(t), \lambda(t)) + b(u^*(t), v, \lambda(t)) \\
 & - (\theta(t) \nabla \mu(t), v) + \langle \varphi_u(u^*(t), \theta^*(t)), v \rangle = 0 \quad \text{with } \lambda(T) = 0 \\
 (3.18) \quad & \left\langle -\frac{d}{dt}\mu(t), \psi \right\rangle + \kappa (\nabla \mu(t), \nabla \psi) - \langle u^*(t) \cdot \nabla \mu(t), \psi \rangle \\
 & - \gamma (\lambda(t), \psi e_d) + \langle \varphi_\theta(u^*(t), \theta^*(t)), \psi \rangle = 0 \quad \text{with } \mu(T) = 0
 \end{aligned}$$

for  $v \in V$  and  $\psi \in H_0^1(\Omega)$  and

$$(3.19) \quad (\beta(g^*(t) - T_0) + \kappa n \cdot \nabla \mu(t), h(t) - g^*(t))_U \geq 0 \quad \text{for all } h \in K.$$

**Proof:** Consider a family of control problems

$$(3.20) \quad \text{minimize } \hat{J}(u, \theta, g) = J(u, \theta, g) + |g - g^*|_U^2 \quad \text{subject to (1.5)}$$

for  $\epsilon > 0$ . It is shown in [IST2] that (3.20) has a solution  $(u_\epsilon^*, \theta_\epsilon^*, g_\epsilon^*)$  where  $u_\epsilon^* \in L^2(0, T; V_0) \cap H^1(0, T; V_0^*)$ ,  $\theta_\epsilon^* \in L^\infty((0, T) \times \Omega)$  and  $g_\epsilon^* \in K$  and they satisfy the necessary optimality condition

$$\begin{aligned}
 & \left\langle -\frac{d}{dt}\lambda_\epsilon(t), v \right\rangle + \nu (\nabla \lambda_\epsilon(t), \nabla v) + b(v, u_\epsilon^*(t), \lambda_\epsilon(t)) + b(u_\epsilon^*(t), v, \lambda_\epsilon(t)) \\
 & - (\theta_\epsilon^*(t) \nabla \mu_\epsilon(t), v) + \langle \varphi_u(u_\epsilon^*(t), \theta_\epsilon^*(t)), v \rangle = 0 \quad \text{with } \lambda_\epsilon(T) = 0 \\
 (3.21) \quad & \left\langle -\frac{d}{dt}\mu_\epsilon(t), \psi \right\rangle + \kappa (\nabla \mu_\epsilon(t), \nabla \psi) + \frac{1}{\epsilon} (\mu_\epsilon(t), \psi)_U - \langle u_\epsilon^*(t) \cdot \nabla \mu_\epsilon(t), \psi \rangle \\
 & - \gamma (\lambda_\epsilon(t), \psi e_d) + \langle \varphi_\theta(u_\epsilon^*(t), \theta_\epsilon^*(t)), \psi \rangle = 0 \quad \text{with } \mu_\epsilon(T) = 0
 \end{aligned}$$

for  $v \in V$  and  $\psi \in H^1(\Omega)$  and

$$(3.22) \quad (\beta(g_\epsilon^*(t) - T_0) + (g_\epsilon^*(t) - g^*(t)) + \kappa n \cdot \nabla \mu_\epsilon(t), h(t) - g_\epsilon^*(t))_U \geq 0 \quad \text{for all } h \in K.$$

Here we used the fact that  $\mu_\epsilon(t) \in L^2(0, T; H^2(\Omega))$  and  $\kappa n \cdot \nabla \mu_\epsilon(t) + \frac{1}{\epsilon} \mu_\epsilon(t) = 0$  on  $\Gamma$ . It follows from the proof of Theorem 3.4 that there exists  $u(t) \in L^2(0, T; V_0) \times H^1(0, T; V_0^*)$ ,  $\theta \in L^\infty((0, T) \times \Omega)$ ,  $g(t) \in K$  and a sequence  $\epsilon$  of positive numbers such that  $\theta_\epsilon^*(t)$  converges to  $\theta(t)$  weakly  $*$  in  $L^\infty((0, T) \times \Omega)$ ,  $u_\epsilon(t)$  converges to  $u(t)$  weakly in  $L^2(0, T; V_0)$  and strongly in  $L^2(0, T; H_0)$ , and  $g_\epsilon^*(t)$  converges to  $g(t)$  weakly in  $U$  where  $(u(t), \theta(t)) \in S(g)$ . Note that

$$J(u_\epsilon^*, \theta_\epsilon^*, g_\epsilon^*) \leq J(u_\epsilon, \theta_\epsilon, g^*),$$

where  $(u_\epsilon, \theta_\epsilon)$  is the solution to (1.5) for  $g(t) = g^*(t)$ . Then, it follows from Corollary 3.5 that  $(u_\epsilon, \theta_\epsilon) \rightarrow (u^*, \theta^*)$ , strongly in  $L^2(0, T; V_0 \times L^2(\Omega))$ . Since  $J$  is lower semicontinuous, we obtain

$$J(u, \theta, g) + |g - g^*|_l^2 \leq J(u^*, \theta^*, g^*)$$

which implies  $g(t) = g^*(t)$  and  $g_\epsilon^*(t) \rightarrow g^*(t)$ , strongly in  $L^2(0, T; L^2(\Gamma))$ .

Now we show that  $(\lambda_\epsilon(t), \mu_\epsilon(t))$  converges to  $(\lambda(t), \mu(t))$ , weakly in  $L^2(0, T; V_0 \times H^2(\Omega))$  and strongly in  $L^2(0, T; H_0 \times H^1(\Omega))$ . Let  $V_2 = \{\psi \in H^2(\Omega) : n \cdot \nabla \psi + \frac{1}{\epsilon} \psi = 0 \text{ on } \Gamma\}$  and consider the sesquilinear form  $\tilde{\sigma}$  on  $V_2 \times V_2$  defined by

$$\tilde{\sigma}(t, \psi_1, \psi_2) = \kappa(\Delta \psi_1, \Delta \psi_2) + (u_\epsilon^*(t) \cdot \nabla \psi_1, \Delta \psi_2)$$

Then,  $\tilde{\sigma}$  satisfies

$$\sup_{\psi_2 \in V_2} |\tilde{\sigma}(t, \psi_1, \psi_2)| \leq \kappa |\Delta \psi_1| + c |u_\epsilon^*(t)|_{H^2} |\nabla \psi_1|_{L^2}$$

and

$$(3.23) \quad \tilde{\sigma}(t, \psi, \psi) \geq \frac{\kappa}{2} |\Delta \psi|_{L^2}^2 - \rho_2(t) |\nabla \psi|_{L^2}^2 \quad \text{for } \psi \in V_2,$$

where  $\rho_2(t) = c |u_\epsilon^*(t)|_{H^2}^2 \in L^1(0, T)$ . Let  $H^1(\Omega)$  be equipped with

$$|\psi|_{H^1}^2 = |\nabla \psi|_{L^2(\Omega)}^2 + \frac{1}{\epsilon} |\psi|_{L^2(\Gamma)}^2.$$

Then,  $L^2(\Omega)$  is the dual space of  $V_2$  when  $H^1(\Omega)$  is identified with its dual. Since for  $\lambda \in V_2$ ,  $\psi \in H^1(\Omega)$

$$\kappa(\nabla \lambda, \nabla \psi) + \frac{1}{\epsilon} (\lambda, \psi)_\Gamma = -\kappa(\Delta \lambda, \psi).$$

The second equation of (3.21) can be written as

$$(3.24) \quad \begin{aligned} & \langle -\frac{d}{dt} \mu_\epsilon(t), \psi \rangle_{L^2 \times V_2} + \tilde{\sigma}(t, \mu_\epsilon(t), \psi) - \gamma(\lambda_\epsilon(t), -\Delta \psi) \\ & + (\varphi_\theta(u_\epsilon^*(t), \theta_\epsilon^*(t)), -\Delta \psi) = 0 \end{aligned}$$

for  $\psi \in V_2$ , where the dual product on  $L^2(\Omega) \times V_2$  is defined as

$$\langle \phi, \psi \rangle_{L^2 \times V_2} = (\phi, -\Delta \psi) \quad \text{for } \phi \in L^2(\Omega), \psi \in V_2.$$

It can be shown, using the Galerkin method in [LM] (see, [IST2] for details), that equation the first equation of (3.21) coupled with (3.24) has a unique solution  $(\lambda_\epsilon(t), \mu_\epsilon(t)) \in L^2(0, T; V_0 \times V_2) \times H^1(0, T; V_0^* \times L^2(\Omega))$ . Moreover, we have

$$\begin{aligned} & \frac{d}{dt} |\mu_\epsilon(t)|_{H^1}^2 + \kappa |\Delta \mu_\epsilon(t)|_{L^2}^2 \\ & \leq 2\rho_2(t) |\mu_\epsilon(t)|_{H^1}^2 + \frac{1}{\kappa} (\gamma |\lambda_\epsilon(t)|_{L^2} + |\varphi_\theta(u_\epsilon^*(t), \theta_\epsilon^*(t))|_{L^2})^2. \end{aligned}$$



Hence, by the Gronwall's inequality  $\sup_{t \in [0, T]} \|\mu_\epsilon(t)\|_{H^1}^2 + \int_0^T \kappa \|\Delta \mu_\epsilon(t)\|_{L^2}^2 dt$  is uniformly bounded in  $0 < \epsilon \leq 1$ , which implies  $\mu_\epsilon \rightarrow 0$  a.e. in  $\Gamma$  uniformly in  $t \in [0, T]$ . Thus, it is not difficult to show that  $(\lambda_\epsilon(t), \mu_\epsilon(t))$  converges to  $(\lambda(t), \mu(t))$ , weakly in  $L^2(0, T; V_0 \times H^2(\Omega))$  and strongly in  $L^2(0, T; H_0 \times H^1(\Omega))$  and  $(\lambda(t), \mu(t))$  satisfies (3.18).  $\square$

## References

- [AT] F. Abergel and R. Temam, On some control problems in fluid mechanics, *Theoretical and Computational Fluid Mechanics*, 1 (1990), 303-325.
- [CF] P. Constantin and C. Foias, *Navier-Stokes Equations*, The University of Chicago Press, Chicago, 1988.
- [DI] M.C. Desai and K. Ito, Optimal Control of Navier-Stokes Equations, *SIAM J. Control & Optim.*, to appear (1994).
- [ET] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North Holland, Amsterdam (1976).
- [Gl] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, Berlin, 1984.
- [Gr] V. Girault and P.A. Raviart, *Finite Element Methods for Navier-Stokes Equations*, Springer-Verlag, Berlin, 1984.
- [IST1] K. Ito, J.S. Seroggs and H.T. Tran, Mathematical issues in optimal design of a vapor transport reactor, *Proceeding of IMA workshop on Flow Control*, ed by M. Gunzburger, (1993).
- [IST2] K. Ito, J.S. Seroggs and H.T. Tran, Optimal Control of Thermally Coupled Navier Stokes Equations, submitted to *SIAM J. Control & Optim.* (1993).
- [LM] J.L. Lions and E. Magenes, *Non-homogeneous Boundary Value Problems and Applications*, Vol I, II, Springer-Verlag, New York, 1972.
- [MZ] H. Maurer and J. Zowe, First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems, *Math Programming*, 16 (1979), 98-110.
- [Ta] H. Tanabe, *Equations of Evolution*, Pitman, San Francisco, 1979.
- [Te1] R. Temam, *Navier-Stokes Equations and Nonlinear Functional Analysis*, SIAM, Philadelphia, 1983.
- [Te2] R. Temam, *Infinite Dimensional Dynamical Systems in Mechanics and Physics*, Appl. Math. Sci. 68, Springer-Verlag, New York, 1988.
- [Tr] G.M. Troianiello, *Elliptic Differential Equations and Obstacle Problems*, Plenum Press, New York, 1987.
- [WvW] Wolf von Wahl, *The Equations of Navier-Stokes and Abstract Parabolic Equations*, Vieweg, Braunschweig 1985.
- [Yo] K. Yosida, *Functional Analysis*, Springer-Verlag, New York.

Kazufumi Ito  
 Center for Research in Scientific Computation  
 Department of Mathematics  
 North Carolina State University  
 Raleigh, NC 27695, USA

## ADAPTIVE ESTIMATION OF NONLINEAR DISTRIBUTED PARAMETER SYSTEMS

JOSEPH KAZIMIR AND I.G. ROSEN

Center for Applied Mathematical Sciences  
Department of Mathematics  
University of Southern California

**ABSTRACT.** The adaptive (on-line) estimation of parameters for a class of nonlinear distributed parameter systems is considered. A combined state and parameter estimator is constructed as an initial value problem for an infinite dimensional evolution equation. State convergence is established via a Lyapunov-like estimate. The finite dimensional notion of persistence of excitation is extended to the infinite dimensional case and used to establish parameter convergence. A finite dimensional approximation theory is presented and a convergence result is proven. An example involving the identification of a nonlinear heat equation is discussed and results of a numerical study are presented.

1991 *Mathematics Subject Classification.* 93B30, 93C25, 93C20, 65J10

*Key words and phrases.* On-line estimation, adaptive identification, parameter convergence, persistence of excitation, distributed parameter systems, infinite dimensional systems, finite dimensional approximation.

### 1. Introduction

In this brief note we consider the adaptive, or on-line, identification of unknown parameters in nonlinear infinite dimensional, or distributed parameter, systems. Our estimator takes the form of an, in general, infinite dimensional nonautonomous *linear* dynamical system. This system is nonautonomous as a result of its dependence upon the state of the plant, which is assumed to be available for measurement in its entirety at all times  $t > 0$ . The estimator is integrated in real time producing estimates for both the plant state and the unknown parameters. The estimator equations are constructed so as to force the derivative of an associated *energy* or Lyapunov functional to be non-positive. It is not, however, negative definite. As a result of this, it is possible to argue, via a modification of a result known as Barbalat's Lemma (a uniformly continuous summable function must tend to zero asymptotically, see [14]), that the state error converges to zero as time tends to infinity. To establish parameter convergence, however, requires an additional *richness* condition on the measured plant data. In the adaptive control literature, this condition is referred to as *persistence of excitation*. We define a

---

This research was supported in part by the Air Force Office of Scientific Research under grant AFOSR 91-0076.

particular notion of richness or persistence of excitation, and then use it to argue that the parameter error converges to zero as time tends to infinity. Finally, since the estimator is infinite dimensional, its integration requires some form of finite dimensional approximation. Consequently, we also develop an associated finite dimensional approximation and convergence theory.

The approach we take here is motivated by the finite dimensional treatments by Morgan and Narendra in [11] and Narendra and Kudva in [12]. We are not the first researchers to suggest that finite dimensional adaptive techniques be applied to infinite dimensional systems. In [1] Alt, Hoffmann, and Sprekels, and in [8] Hoffmann and Sprekels considered an asymptotic embedding method for the identification of linear elliptic partial differential equations. The elliptic equation is embedded in an evolution equation so that the elliptic equation's solution is an asymptotic steady state of the evolution equation. In [9] (see also [7]) Hoffmann and Sprekels provide an abstract functional analytic framework for their earlier results and extend them to certain classes of stationary elliptic and evolutionary parabolic nonlinear variational inequalities. Our treatment here is most closely related to the work of Baumeister and Scordo in [2] (see also [3]) and the results in Scordo's Ph.D. thesis [16]. The primary difference between these efforts and ours is that theirs is a *two-space* theory, while ours is set in only a single Hilbert space. More precisely, their state estimator is defined via a Gelfand triple of spaces (i.e.  $V \hookrightarrow H \hookrightarrow V^*$ ) and a strongly  $V$ -coercive operator. Our state estimator, on the other hand, is set in the same Hilbert space,  $H$ , as the plant and is governed by an  $H$ -accretive operator. To establish convergence in our case, we had to appropriately modify the requisite regularity and richness conditions on the plant. In our definitions of admissibility and persistence of excitation, the interaction between the plant and the estimator dynamics is more explicit than it is in the two-space theory. Interaction between the plant and estimator dynamics in the two-space theory is certainly present. However, it is somewhat subtler. The one-space theory is in some ways more versatile than the corresponding two-space theory. This is particularly true in the case of hereditary or delay systems where in general there is no natural and simple choice for an energy space.

This paper represents our first reporting of our findings and is by no means complete. For example, a careful examination and characterization of the persistence of excitation condition given in Section 3 in the spirit of the one given for the two-space theory in [6] must be carried out. The same is true for the admissibility condition given in Section 2. The significance of such studies stems from the fact that in actual practice, with the possible exception of only the simplest plants and estimators, these conditions are extremely difficult, if not impossible, to check. Our efforts in this direction and others are currently underway. When these studies have been completed, our findings will be reported on elsewhere.

An outline of the remainder of the paper is as follows. In Section 2 we define the plant and the estimator. The notion of an *admissible* plant is defined and the estimator equations are shown to be well posed. In Section 3 the convergence of the state error to zero is established. We define the notion of a plant being

persistently exciting, and use this to establish parameter convergence. In Section 4 we develop our abstract approximation theory, and provide sufficient conditions for an approximation scheme to be convergent. Finally, in a fifth section, to illustrate the application of our general approach, we consider the identification of a gradient dependent thermal conductivity in a one dimensional nonlinear heat equation with Dirichlet boundary conditions. An estimator in the form of a simple constant coefficient heat operator with Neumann boundary conditions is defined. A linear spline based approximation scheme is developed. Numerical results are presented.

## 2. An Adaptive or On-Line Estimator

Let  $H$  be a real Hilbert space with inner product  $\langle \cdot, \cdot \rangle$  and corresponding induced norm  $|\cdot|$ , and let  $Q$  be a real Hilbert space with inner product  $\langle \cdot, \cdot \rangle_Q$  and corresponding induced norm  $|\cdot|_Q$ . For each  $q \in Q$  let  $A(q) : \text{Dom}(A(q)) \subset H \rightarrow H$  be a possibly nonlinear operator on  $H$ , where  $\text{Dom}(A(q)) = \{\varphi \in H : A(q)\varphi \in H\}$ . We assume that the map  $q \rightarrow A(q)\varphi$ ,  $\varphi \in \text{Dom}(A(q))$  is linear from  $Q$  into  $H$ . Let  $\bar{f} : [0, \infty) \rightarrow H$ , and let  $\bar{x}_0 \in H$ . Let  $\bar{q} \in Q$  and consider the dynamical system in  $H$  given by

$$(2.1) \quad \dot{\bar{x}}(t) + A(\bar{q})\bar{x}(t) = \bar{f}(t), \quad t > 0,$$

$$(2.2) \quad \bar{x}(0) = \bar{x}_0.$$

We assume that  $\bar{q}$  is unknown and that the state of the plant, (2.1), (2.2),  $\bar{x}(t)$  can be measured and is available at all times  $t > 0$ . We want to define a linear dynamical system on  $H$  which uses the measurement of  $\bar{x}$  to asymptotically (i.e. as  $t \rightarrow \infty$ ) estimate  $\bar{q}$ .

Let  $A_0 : \text{Dom}(A_0) \subset H \rightarrow H$  be a linear operator on  $H$  satisfying the following two assumptions.

(A1) (Accretivity) There exists  $\alpha_0 \geq 0$  for which

$$\langle A_0\varphi, \varphi \rangle \geq \alpha_0|\varphi|^2, \quad \varphi \in \text{Dom}(A_0).$$

(A2) (Surjectivity) The operator  $\lambda + A_0 : \text{Dom}(A_0) \subset H \rightarrow H$  is surjective for some  $\lambda > 0$ .

It follows (see, for example, [10], [13] or [17]) that  $\overline{\text{Dom}(A_0)} = H$ , and that  $-A_0 \in G(H, 1, -\alpha_0)$ . That is,  $-A_0$  is the infinitesimal generator of a  $C_0$ -semigroup of bounded linear operators on  $H$ ,  $\{T_0(t) : t \geq 0\}$ , satisfying

$$(2.3) \quad |T_0(t)| \leq e^{-\alpha_0 t}, \quad t \geq 0.$$

Our on-line estimator is then defined to be the dynamical system in  $H$  given by

$$(2.4) \quad \dot{x}(t) + A_0x(t) + A(q(t))\bar{x}(t) = \bar{f}(t) + A_0\bar{x}(t), \quad t > 0,$$

$$(2.5) \quad \langle \dot{q}(t), p \rangle_Q - \langle A(p)\bar{x}(t), x(t) \rangle = -\langle A(p)\bar{x}(t), \bar{x}(t) \rangle, \quad p \in Q, \quad t > 0,$$

$$(2.6) \quad x(0) = x_0 \in H, \quad q(0) = q_0 \in Q.$$

The definition of our estimator given in (2.4)-(2.6) above is primarily motivated by the analogous definition in finite dimensions (see [12]). In the finite dimensional setting, the construction of the estimator in this form is based upon the desire for linearity, homogeneous error equations, and the ability to obtain a particular Lyapunov estimate. This estimator can also be derived via a gradient argument (see [16]).

We assume that  $\bar{x}$  is a strong solution to the initial value problem (2.1), (2.2) in the sense that  $\bar{x} \in C^1([0, \infty); H)$  and  $\bar{x}(t) \in \text{Dom}(A(\bar{q}))$  for all  $t \geq 0$ . We make the additional regularity assumption on the data,  $\bar{x}$ , that for all  $t \geq 0$ ,  $\bar{x}(t) \in \text{Dom}(A(q))$ , for all  $q \in Q$ . We note that if the operator  $A$  and the parameter space  $Q$  are sufficiently regular in the sense that  $\text{Dom}(A(q)) = \text{Dom}(A)$  is independent of  $q \in Q$ , then this additional regularity assumption on the data  $\bar{x}$  can be eliminated.

For  $t \geq 0$ , define the linear map  $B(t) : Q \rightarrow H$  by

$$B(t)q = A(q)\bar{x}(t), \quad q \in Q.$$

To establish that the estimator (2.4)-(2.6) is well posed, and to establish our convergence results in the next section, we require the following definition.

**Definition 2.1.** The triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  will be called  $A_0$ -admissible if

- (i) for each  $t \geq 0$ , the operator  $B(t)$  is bounded (i.e.,  $B(t) \in \mathcal{L}(Q, H)$ ) and there exists  $\alpha > 0$  such that

$$|B(t)q| \leq \alpha |q|_Q, \quad q \in Q,$$

- (ii) the operators  $B(t)^* \in \mathcal{L}(H, Q)$ ,  $t \geq 0$ , are strongly continuously differentiable on  $[0, \infty)$ ,
- (iii)  $\bar{x}(t) \in \text{Dom}(A_0)$ , for  $t \geq 0$ , and  $A_0\bar{x} \in C^1([0, \infty); H)$ ,
- (iv)  $\bar{f} \in C^1([0, \infty); H)$ .

We note that when  $B(t) \in \mathcal{L}(Q, H)$ , its Hilbert space adjoint,  $B(t)^* \in \mathcal{L}(H, Q)$  is given by

$$\langle B(t)^*\varphi, q \rangle_Q = \langle A(q)\bar{x}(t), \varphi \rangle_H, \quad \varphi \in H, q \in Q.$$

Let  $Z = H \times Q$  with inner product  $\langle \cdot, \cdot \rangle_Z$  given by

$$(2.7) \quad \langle (\varphi, q), (\psi, p) \rangle_Z = \langle \varphi, \psi \rangle + \langle q, p \rangle_Q, \quad (\varphi, q), (\psi, p) \in Z,$$

with the corresponding induced norm being denoted by  $|\cdot|_Z$ . For  $t \geq 0$ , define the linear operator  $\mathcal{A}(t) : \text{Dom}(\mathcal{A}) \subset Z \rightarrow Z$  by

$$\mathcal{A}(t) = \begin{bmatrix} A_0 & B(t) \\ -B(t)^* & 0 \end{bmatrix}$$

with  $\text{Dom}(\mathcal{A}) = \text{Dom}(A_0) \times Q$ , define  $F(t) \in Z$  by

$$F(t) = \begin{bmatrix} \bar{f}(t) + A_0 \bar{x}(t) \\ -B(t)^* \bar{x}(t) \end{bmatrix},$$

and set  $z_0 = (x_0, q_0) \in Z$ . Note that  $\text{Dom}(\mathcal{A})$  is dense in  $Z$ . The estimator (2.4)-(2.6) can then be written as the abstract initial value problem in  $Z$  given by

$$(2.8) \quad \dot{z}(t) + \mathcal{A}(t)z(t) = F(t), \quad t > 0,$$

$$(2.9) \quad z(0) = z_0.$$

Assume that the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible, and define the operator  $\mathcal{A}_0 : \text{Dom}(\mathcal{A}) \subset Z \rightarrow Z$  by

$$\mathcal{A}_0 = \begin{bmatrix} A_0 & 0 \\ 0 & 0 \end{bmatrix},$$

and for  $t \geq 0$  define the operator  $\mathcal{B}(t) \in \mathcal{L}(Z, Z)$  by

$$\mathcal{B}(t) = \begin{bmatrix} 0 & B(t) \\ -B(t)^* & 0 \end{bmatrix}.$$

Then  $\mathcal{A}(t) = \mathcal{A}_0 + \mathcal{B}(t)$ ,  $t \geq 0$ , and it is easily argued that for  $t \geq 0$

$$|\langle \mathcal{B}(t)z_1, z_2 \rangle_Z| \leq \alpha |z_1|_Z |z_2|_Z, \quad z_1, z_2 \in Z,$$

and thus that  $\|\mathcal{B}(t)\|_{\mathcal{L}(Z, Z)} \leq \alpha$ ,  $t \geq 0$ . The fact that  $-A_0 \in G(H, 1, -\alpha_0)$  implies that  $-\mathcal{A}_0$  is the infinitesimal generator of a  $C_0$ -semigroup of contractions,  $\{\mathcal{T}_0(t) : t \geq 0\}$ , on  $Z$ . That is,  $-\mathcal{A}_0 \in G(Z, 1, 0)$ . It follows that (see, for example, [10], Theorem IX.2.1) for each  $t \geq 0$ ,  $-\mathcal{A}(t) \in G(Z, 1, \alpha)$ . Consequently, the family of operators  $\{-\mathcal{A}(t) : 0 \leq t \leq T\}$  is *stable* (see [18], Definition 4.3) for all  $T > 0$ .

The fact that it was assumed that  $\bar{x}$  is a strong solution to the plant equations, (2.1), (2.2), and that the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible, implies that the map  $t \rightarrow -\mathcal{A}(t)z$  is strongly continuously differentiable in  $Z$  for each  $z \in \text{Dom}(\mathcal{A})$ . It follows from the corollary to Theorem 4.4.2 in [18] that there exists a unique evolution system,  $\{U(t, s) : 0 \leq s \leq t\}$ , with  $U(t, s) \in \mathcal{L}(Z, Z)$ ,  $0 \leq s \leq t$  satisfying

- (a)  $U(t, s)$  is strongly continuous in  $s$  and  $t$ ,  $U(s, s) = I$ , and  $\|U(t, s)\| \leq e^{\alpha(t-s)}$ ,
- (b)  $U(t, s) = U(t, r)U(r, s)$ , for  $0 \leq s \leq r \leq t$ ,
- (c)  $U(t, s)\text{Dom}(\mathcal{A}) \subset \text{Dom}(\mathcal{A})$ , for  $0 \leq s \leq t$ ,
- (d) for  $z \in \text{Dom}(\mathcal{A})$ ,  $U(t, s)z$  is strongly continuously differentiable on  $Z$  in both  $s$  and  $t$ , and

$$\frac{\partial}{\partial t} U(t, s)z = -\mathcal{A}(t)U(t, s)z, \quad 0 < s < t,$$

$$\frac{\partial}{\partial s} U(t, s)z = U(t, s)\mathcal{A}(s)z, \quad 0 < s < t,$$

with both sides of these equations being strongly continuous in  $Z$ .

Continuing to assume that the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible, define the function  $z \in C^1([0, T]; Z)$ , for all  $T > 0$  by

$$(2.10) \quad z(t) = U(t, 0)z_0 + \int_0^t U(t, s)F(s)ds, \quad t \geq 0.$$

The following theorem and its corollary are then immediate consequences of Theorem 4.5.3 in [18].

**Theorem 2.2.** *Suppose that the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible and that  $z_0 = (x_0, q_0) \in \text{Dom}(\mathcal{A})$ . Then the  $Z$ -valued function  $z$  given by (2.10) is an element in  $C^1([0, T]; Z)$ , for all  $T > 0$ , and satisfies (2.8) and (2.9).*

**Corollary 2.3.** *Suppose that the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible, that  $q_0 \in Q$  and  $x_0 \in \text{Dom}(A_0)$ , and for  $t \geq 0$ , we write  $z(t)$  given by (2.10) as  $z(t) = (x(t), q(t))$ . Then  $x \in C^1([0, T]; H)$  and  $q \in C^1([0, T]; Q)$ , for all  $T > 0$ , and the pair  $(x, q)$  satisfies (2.4)-(2.6).*

For  $t \geq 0$  set  $e(t) = x(t) - \bar{x}(t)$  and  $r(t) = q(t) - \bar{q}$ . It then follows that  $e$  and  $r$  satisfy

$$(2.11) \quad \dot{e}(t) + A_0 e(t) + A(r(t))\bar{x}(t) = 0, \quad t > 0,$$

$$(2.12) \quad \langle \dot{e}(t), p \rangle_Q - \langle A(p)\bar{x}(t), e(t) \rangle = 0, \quad p \in Q, \quad t > 0,$$

$$(2.13) \quad e(0) = e_0 \in H, \quad r(0) = r_0 \in Q,$$

where  $e_0 = x_0 - \bar{x}_0$  and  $r_0 = q_0 - \bar{q}$ . The system (2.11)-(2.13) may equivalently be written

$$y(t) + \mathcal{A}(t)y(t) = 0, \quad t > 0,$$

$$y(0) = y_0$$

where for  $t \geq 0$ ,  $y(t) = (e(t), r(t)) \in Z$  and  $y_0 = (e_0, r_0) \in Z$ . It follows that

$$y(t) = U(t, s)y(s), \quad 0 \leq s \leq t.$$

If the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible and  $e_0 \in \text{Dom}(A_0)$  and  $r_0 \in Q$ , then  $y \in C^1([0, T]; Z)$ ,  $e \in C^1([0, T]; H)$ , and  $r \in C^1([0, T]; Q)$ , for all  $T > 0$ . In the next section we argue that  $\lim_{t \rightarrow \infty} |e(t)| = 0$  and, under an appropriate additional richness assumption on the plant data,  $\bar{x}(t)$ ,  $t \geq 0$ , (i.e.  $A_0$ -persistence of excitation) that  $\lim_{t \rightarrow \infty} |r(t)|_Q = 0$ .

### 3. State and Parameter Error Convergence

Let  $e$  and  $r$  be the solutions to the initial value problem (2.11)-(2.13) and define

$$E(t) = \frac{1}{2} \{ |e(t)|^2 + |r(t)|_Q^2 \}, \quad t \geq 0.$$

**Lemma 3.1.** Suppose that the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible and that  $x_0 \in \text{Dom}(A_0)$ . Then

$$(3.1) \quad E(t) + \alpha_0 \int_0^t |e(s)|^2 ds \leq \zeta, \quad t \geq 0,$$

where  $\zeta = E(0) = \frac{1}{2} \{ |e_0|^2 + |r_0|_Q^2 \}$ .

**Proof.** Using (2.11) and (2.12) we obtain

$$\begin{aligned} \frac{d}{dt} E(t) &= \frac{d}{dt} \frac{1}{2} \{ \langle e(t), e(t) \rangle + \langle r(t), r(t) \rangle_Q \} \\ &= \langle \dot{e}(t), e(t) \rangle + \langle \dot{r}(t), r(t) \rangle_Q \\ (3.2) \quad &= -\langle A_0 e(t) + B(t)r(t), e(t) \rangle + \langle B(t)^* e(t), r(t) \rangle_Q \\ &= -\langle A_0 e(t), e(t) \rangle \\ &\leq -\alpha_0 |e(t)|^2. \end{aligned}$$

Integrating both sides of the above expression from 0 to  $t$ , we obtain (3.1).  $\square$

The convergence of the state estimator is established in the next theorem. The proof we give is a variant of the argument used to establish Barbalat's Lemma in [14].

**Theorem 3.2.** If the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible and  $x_0 \in \text{Dom}(A_0)$ , then  $E$  is nonincreasing and  $\lim_{t \rightarrow \infty} |e(t)| = 0$ .

**Proof.** That  $E$  is nonincreasing follows immediately from (3.2). Let  $t_2 \geq t_1 \geq 0$  and note that Assumption (A1), Definition 2.1, and (3.1) imply that

$$\begin{aligned} |e(t_2)|^2 - |e(t_1)|^2 &= \int_{t_1}^{t_2} \frac{d}{dt} |e(t)|^2 dt \\ &= 2 \int_{t_1}^{t_2} \langle \dot{e}(t), e(t) \rangle dt \\ (3.3) \quad &= -2 \int_{t_1}^{t_2} \langle A_0 e(t), e(t) \rangle dt - 2 \int_{t_1}^{t_2} \langle B(t)r(t), e(t) \rangle dt \\ &\leq -2\alpha_0 \int_{t_1}^{t_2} |e(t)|^2 dt + 2\alpha \int_{t_1}^{t_2} |r(t)|_Q |e(t)| dt \\ &\leq 2\alpha \int_{t_1}^{t_2} \frac{1}{2} \{ |e(t)|^2 + |r(t)|_Q^2 \} dt \\ &\leq 2\alpha \zeta (t_2 - t_1). \end{aligned}$$



Now suppose that  $\lim_{t \rightarrow \infty} |c(t)| \neq 0$ . Then there exist  $\varepsilon > 0$  and a sequence  $\{t_i\}_{i=1}^\infty$  with  $\lim_{i \rightarrow \infty} t_i = \infty$  for which

$$(3.4) \quad |c(t_i)|^2 > \varepsilon, \quad i = 1, 2, \dots$$

It follows from (3.3) and (3.4) that for  $\delta > 0$  and  $i = 1, 2, \dots$

$$\begin{aligned} \int_{t_i - \delta}^{t_i} |c(t)|^2 dt &= \int_{t_i - \delta}^{t_i} |c(t_i)|^2 dt - \int_{t_i - \delta}^{t_i} \{|c(t_i)|^2 - |c(t)|^2\} dt \\ &> \varepsilon \delta - 2\alpha \zeta \int_{t_i - \delta}^{t_i} (t_i - t) dt \\ &= \varepsilon \delta - \alpha \zeta \delta^2. \end{aligned}$$

Choosing  $\delta = \varepsilon/(2\alpha\zeta)$ , we obtain

$$\int_{t_i - \delta}^{t_i} |c(t)|^2 dt \geq \frac{\varepsilon \delta}{2}, \quad i = 1, 2, \dots$$

This contradicts the fact that Lemma 3.1 implies that

$$\alpha_0 \int_0^\infty |c(t)|^2 dt \leq \zeta < \infty,$$

and consequently  $\lim_{t \rightarrow \infty} |c(t)|^2 = 0$ , and the theorem is proved.  $\square$

In order to establish parameter convergence, we require the following definition.

**Definition 3.3.** The triple  $\{\bar{q}, \bar{\tau}_0, \bar{f}\}$  is said to be either  $A_0$ -persistently exciting or  $A_0$ -rich if there exist positive constants  $\tau_0$ ,  $\delta_0$ , and  $\varepsilon_0$  such that for each  $q \in Q$  with  $|q|_Q = 1$  and  $t \geq 0$  sufficiently large, there exists  $\bar{t} \in [t, t + \tau_0]$  such that

$$\left| \int_{\bar{t}}^{\bar{t} + \delta_0} T_0(\bar{t} + \delta_0 - s) B(s) q ds \right| \geq \varepsilon_0.$$

Parameter convergence is demonstrated in the next theorem. The proof is a variant on an argument used in the finite dimensional case by Baumeister and Scondo in [2].

**Theorem 3.4.** Suppose that the triple  $\{\bar{q}, \bar{\tau}_0, \bar{f}\}$  is  $A_0$ -admissible and  $A_0$ -persistently exciting and that  $x_0 \in \text{Dom}(A_0)$ . Then  $\lim_{t \rightarrow \infty} |r(t)|_Q = 0$ .

**Proof.** We begin by noting that for  $t_2 \geq t_1$ , (2.11) implies that

$$c(t_2) = T_0(t_2 - t_1)c(t_1) - \int_{t_1}^{t_2} T_0(t_2 - s)B(s)r(s)ds.$$

It then follows from (2.3) that

$$\begin{aligned} \left| \int_{t_1}^{t_2} T_0(t_2 - s) B(s) r(s) ds \right| &\leq |c(t_2)| + |T_0(t_2 - t_1) c(t_1)| \\ (3.5) \qquad \qquad \qquad &\leq |c(t_2)| + e^{-\alpha_0(t_2 - t_1)} |c(t_1)|. \end{aligned}$$

Also, from (2.12) and Definition 2.1 we obtain

$$\begin{aligned} |r(t_2) - r(t_1)|_Q &= \sup_{|q|_Q \leq 1} |\langle r(t_2) - r(t_1), q \rangle_Q| = \sup_{|q|_Q \leq 1} |\langle \int_{t_1}^{t_2} \dot{r}(t) dt, q \rangle_Q| \\ (3.6) \qquad \qquad &= \sup_{|q|_Q \leq 1} \left| \int_{t_1}^{t_2} \langle \dot{r}(t), q \rangle_Q dt \right| = \sup_{|q|_Q \leq 1} \left| \int_{t_1}^{t_2} \langle B(t)^* c(t), q \rangle_Q dt \right| \\ &= \sup_{|q|_Q \leq 1} \left| \int_{t_1}^{t_2} \langle A(q) \bar{r}(t), c(t) \rangle dt \right| \leq \sup_{|q|_Q \leq 1} \int_{t_1}^{t_2} \alpha |q|_Q |c(t)| dt \\ &\leq \alpha \int_{t_1}^{t_2} |c(t)| dt \leq \alpha(t_2 - t_1)^{\frac{1}{2}} \left\{ \int_{t_1}^{t_2} |c(t)|^2 dt \right\}^{\frac{1}{2}}. \end{aligned}$$

Now, from Theorem 3.2 we know that  $E$  is nonincreasing. Since  $E(t) \geq 0$ ,  $t \geq 0$ , it follows that  $\lim_{t \rightarrow \infty} E(t)$  must exist. Also from Theorem 3.2 we know that  $\lim_{t \rightarrow \infty} |c(t)| = 0$ . Consequently, we must have that  $\lim_{t \rightarrow \infty} |r(t)|_Q = \rho$ , for some  $\rho \geq 0$ . Suppose that  $\rho \neq 0$ . Then there exist a sequence of positive numbers  $\{t_k\}_{k=1}^{\infty}$  with  $\lim_{k \rightarrow \infty} t_k = \infty$  and  $\delta > 0$  for which  $|r(t_k)|_Q > \delta$ ,  $k = 1, 2, \dots$ . For each positive integer  $k$  sufficiently large, let  $\tilde{t}_k$  be as in Definition 3.3 corresponding to  $q = r(t_k)/|r(t_k)|_Q$ . Then, using (3.5) and (3.6) we obtain

$$\begin{aligned} &\left| \int_{\tilde{t}_k}^{\tilde{t}_k + \delta_0} T_0(\tilde{t}_k + \delta_0 - t) B(t) r(t) dt \right| \\ &\leq \left| \int_{\tilde{t}_k}^{\tilde{t}_k + \delta_0} T_0(\tilde{t}_k + \delta_0 - t) B(t) r(t) dt \right| \\ &+ \left| \int_{\tilde{t}_k}^{\tilde{t}_k + \delta_0} T_0(\tilde{t}_k + \delta_0 - t) B(t) (r(t_k) - r(t)) dt \right| \\ &\leq |c(\tilde{t}_k + \delta_0)| + e^{-\alpha_0 \delta_0} |c(\tilde{t}_k)| \\ &+ \left| \int_{\tilde{t}_k}^{\tilde{t}_k + \delta_0} e^{-\alpha_0(\tilde{t}_k + \delta_0 - t)} \alpha^2 (t - t_k)^{\frac{1}{2}} \left\{ \int_{t_k}^t |c(s)|^2 ds \right\}^{\frac{1}{2}} dt \right| \\ &\leq |c(\tilde{t}_k + \delta_0)| + e^{-\alpha_0 \delta_0} |c(\tilde{t}_k)| + \alpha^2 (\tilde{t}_k + \delta_0 - t_k)^{\frac{1}{2}} \left\{ \int_{t_k}^{\tilde{t}_k + \delta_0} |c(s)|^2 ds \right\}^{\frac{1}{2}} \delta_0 \\ &\leq |c(\tilde{t}_k + \delta_0)| + e^{-\alpha_0 \delta_0} |c(\tilde{t}_k)| + \alpha^2 \delta_0 \sqrt{\tau_0 + \delta_0} \left\{ \int_{t_k}^{t_k + \tau_0 + \delta_0} |c(s)|^2 ds \right\}^{\frac{1}{2}}. \end{aligned}$$

It then follows from Definition 3.3 and the above estimate that for all positive integers  $k$  sufficiently large, we have

$$\begin{aligned} 0 < \varepsilon_0 \delta &\leq |r(t_k)|_Q \int_{\tilde{t}_k}^{\tilde{t}_k + \delta_0} T_0(\tilde{t}_k + \delta_0 - t) B(t) \frac{r(t_k)}{|r(t_k)|_Q} dt \\ (3.7) \quad &\leq |e(\tilde{t}_k + \delta_0)| + e^{-\alpha_0 \delta_0} |e(\tilde{t}_k)| + \alpha^2 \delta_0 \sqrt{\tau_0 + \delta_0} \left\{ \int_{\tilde{t}_k}^{\tilde{t}_k + \tau_0 + \delta_0} |e(t)|^2 dt \right\}^{\frac{1}{2}}. \end{aligned}$$

Letting  $k \rightarrow \infty$  in (3.7), and recalling Lemma 3.1 and Theorem 3.2, we obtain a contradiction. Thus we must have that  $\lim_{t \rightarrow \infty} |r(t)|_Q = 0$ , and the theorem is proved.  $\square$

#### 4. Approximation Theory

The estimator, (2.4)-(2.6) is, in general, infinite dimensional. Consequently its implementation requires some form of finite dimensional approximation and a corresponding convergence theory.

For each  $n = 1, 2, \dots$ , let  $H^n$  and  $Q^n$  be finite dimensional subspaces of  $H$  and  $Q$ , respectively. Let  $P_H^n$  and  $P_Q^n$  denote the corresponding orthogonal projections of  $H$  onto  $H^n$  and  $Q$  onto  $Q^n$ , respectively. Let  $A_0^n \in \mathcal{L}(H^n, H^n)$ . In order to establish a convergence result, we will require the following three assumptions.

(A3) (Subspace Approximation) For each  $\varphi \in H$  and  $q \in Q$  we have

$$\lim_{n \rightarrow \infty} |P_H^n \varphi - \varphi| = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} |P_Q^n q - q|_Q = 0.$$

(A4) (Uniform Accretivity) For  $\alpha_0 \geq 0$  as in Assumption (A1), we have

$$\langle A_0^n \varphi^n, \varphi^n \rangle \geq \alpha_0 |\varphi^n|^2, \quad \varphi^n \in H^n.$$

(A5) (Strong Convergence in the Generalized Sense) For some  $\lambda$  with  $\operatorname{Re} \lambda > -\alpha_0$ , we have

$$\lim_{n \rightarrow \infty} |(\lambda + A_0^n)^{-1} P_H^n \varphi - (\lambda + A_0)^{-1} \varphi| = 0, \quad \varphi \in H.$$

Let  $\{T_0^n(t) : t \geq 0\}$  denote the  $C_0$ -semigroup of bounded linear operators on  $H^n$  with infinitesimal generator  $-A_0^n$ . That is,

$$T_0^n(t) = \exp(-A_0^n t), \quad t \geq 0.$$

It follows from Assumption (A4) above that

$$|T_0^n(t)| \leq e^{-\alpha_0 t}, \quad t \geq 0,$$

and from Assumptions (A3) - (A5) and the Trotter-Kato theorem (see, for example, [10] or [13]) that

$$(4.1) \quad \lim_{n \rightarrow \infty} |T_0^n(t) P_H^n \varphi - T_0(t) \varphi| = 0, \quad \varphi \in H,$$

with the convergence being uniform in  $t$  on compact  $t$ -intervals.

For each  $t \geq 0$  define  $B^n(t) \in \mathcal{L}(Q^n, H^n)$  by

$$B^n(t)q^n = P_H^n B(t)q^n = P_H^n A(q^n)\bar{x}(t), \quad q^n \in Q^n.$$

If the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible, then  $|B^n(t)| \leq \alpha$ ,  $t \geq 0$ .

We consider the following approximating estimator,

$$(4.2) \quad \dot{x}^n(t) + A_0^n x^n(t) + B^n(t)q^n(t) = P_H^n \bar{f}(t) + P_H^n A_0 \bar{x}(t), \quad t > 0,$$

$$(4.3) \quad \dot{q}^n(t) - B^n(t)^* x^n(t) = -P_Q^n B(t)^* \bar{x}(t), \quad t > 0,$$

$$(4.4) \quad x^n(0) = x_0^n = P_H^n x_0, \quad q^n(0) = q_0^n = P_Q^n q_0,$$

and argue that

$$(4.5) \quad \lim_{n \rightarrow \infty} |x^n(t) - x(t)| = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} |q^n(t) - q(t)|_Q = 0,$$

uniformly in  $t$  on compact  $t$ -intervals, where  $x^n$  and  $q^n$  are the solutions to the system (4.2)-(4.4) and  $x$  and  $q$  are the solutions to the system (2.4)-(2.6).

To establish the convergence in (4.5), we rewrite the initial value problem (4.2)-(4.4) as an equivalent system in the product space  $Z^n = H^n \times Q^n$ . Note that  $Z^n$  is a finite dimensional subspace of  $Z$ . For  $t \geq 0$ , define the linear operator  $\mathcal{A}^n(t) \in \mathcal{L}(Z^n, Z^n)$  by

$$\mathcal{A}^n(t) = \begin{bmatrix} A_0^n & B^n(t) \\ -B^n(t)^* & 0 \end{bmatrix}.$$

For each  $n = 1, 2, \dots$ ,  $\mathcal{L}(Z, Z^n)$  denote the orthogonal projection of  $Z$  onto  $Z^n$  with respect to the inner product given in (2.7). It follows that

$$P^n = \begin{bmatrix} P_H^n & 0 \\ 0 & P_Q^n \end{bmatrix}.$$

For each  $t \geq 0$  set  $F^n(t) = P^n F(t) \in Z^n$ ; that is

$$F^n(t) = \begin{bmatrix} P_H^n \bar{f}(t) + P_H^n A_0 \bar{x}(t) \\ -P_Q^n B(t)^* \bar{x}(t) \end{bmatrix}.$$

We set  $z_0^n = (x_0^n, q_0^n) \in Z^n$ . The estimator (4.2)-(4.4) can then be written as the initial value problem in  $Z^n$  given by

$$(4.6) \quad \dot{z}^n(t) + \mathcal{A}^n(t)z^n(t) = F^n(t), \quad t > 0,$$

$$(4.7) \quad z^n(0) = z_0^n,$$

where  $z^n(t) = (x^n(t), q^n(t))$ ,  $t \geq 0$ .

Assume that the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible, and let  $\{U^n(t, s) : 0 \leq s \leq t\}$  be the fundamental operator solution to the initial value problem (4.6), (4.7). It then follows that

$$|U^n(t, s)| \leq e^{\alpha(t-s)}, \quad 0 \leq s \leq t,$$

and that

$$(4.8) \quad z^n(t) = U^n(t, 0)z_0^n + \int_0^t U^n(t, s)F^n(s)ds, \quad t \geq 0,$$

and we want to show that

$$\lim_{n \rightarrow \infty} |z^n(t) - z(t)|_Z = 0,$$

uniformly in  $t$  on compact  $t$ -intervals, where  $z^n$  is given by (4.8) and  $z$  is given by (2.10).

Define the operator  $\mathcal{A}_0^n \in \mathcal{L}(Z^n, Z^n)$  by

$$\mathcal{A}_0^n = \begin{bmatrix} A_0^n & 0 \\ 0 & 0 \end{bmatrix},$$

and for  $t \geq 0$  define the operator  $\mathcal{B}^n(t) \in \mathcal{L}(Z^n, Z^n)$  by

$$\mathcal{B}^n(t) = \begin{bmatrix} 0 & B^n(t) \\ -B^n(t)^* & 0 \end{bmatrix}.$$

Then  $\mathcal{A}^n(t) = \mathcal{A}_0^n + \mathcal{B}^n(t)$ ,  $t \geq 0$ ,  $\mathcal{B}^n(t) = \mathcal{P}^n \mathcal{B}(t)$ ,  $t \geq 0$ , and it is easily argued that for  $t \geq 0$ ,  $|\mathcal{B}^n(t)| \leq \alpha$ . Let  $\{\mathcal{T}_0^n(t) : t \geq 0\}$  be the  $C_0$ -semigroup of contractions on  $Z^n$  with infinitesimal generator  $-\mathcal{A}_0^n$ . That is  $\mathcal{T}_0^n(t) = \exp(-\mathcal{A}_0^n t)$ ,  $t \geq 0$ , with  $|\mathcal{T}_0^n(t)| \leq 1$ ,  $t \geq 0$ . It also follows that

$$\mathcal{T}_0^n(t) = \begin{bmatrix} T_0^n(t) & 0 \\ 0 & I \end{bmatrix}, \quad t \geq 0,$$

and from (4.1), that

$$(4.9) \quad \lim_{n \rightarrow \infty} |\mathcal{T}_0^n(t) \mathcal{P}^n \varphi - \mathcal{T}_0(t) \varphi|_Z = 0, \quad \varphi \in Z,$$

uniformly in  $t$  on compact  $t$ -intervals, where  $\{\mathcal{T}_0(t) : t \geq 0\}$  is the  $C_0$ -semigroup of contractions on  $Z$  with generator  $-\mathcal{A}_0$ .

**Theorem 4.1.** Suppose that the Assumptions (A1)-(A5) hold and that the triple  $\{\bar{q}, \bar{x}_0, \bar{f}\}$  is  $A_0$ -admissible. Then

$$\lim_{n \rightarrow \infty} |z^n(t) - z(t)|_Z = 0,$$

uniformly in  $t$  on compact  $t$ -intervals, where  $z^n$  is given by (4.8) and  $z$  is given by (2.10).

**Proof.** From (2.10) and (4.8), it is clear that we need only argue that

$$(4.10) \quad \lim_{n \rightarrow \infty} \|U^n(t, s)\mathcal{P}^n \varphi - U(t, s)\varphi\|_Z = 0, \quad \varphi \in Z,$$

uniformly on the triangle  $\Delta_T = \{(s, t) : 0 \leq s \leq t \leq T\}$ , where  $T > 0$  is fixed but arbitrary. Now, for  $(s, t) \in \Delta_T$  we have (see [4], Theorem 5.2)

$$U^n(t, s)\mathcal{P}^n \varphi = \mathcal{T}_0^n(t-s)\mathcal{P}^n \varphi + \int_s^t \mathcal{T}_0^n(t-r)\mathcal{B}^n(r)U^n(r, s)\mathcal{P}^n \varphi dr, \quad \varphi \in Z,$$

and

$$U(t, s)\varphi = \mathcal{T}_0(t-s)\varphi + \int_s^t \mathcal{T}_0(t-r)\mathcal{B}(r)U(r, s)\varphi dr, \quad \varphi \in \text{Dom}(\mathcal{A}).$$

Consequently, for  $\varphi \in \text{Dom}(\mathcal{A})$  and  $(s, t) \in \Delta_T$ , we have

$$\begin{aligned} & \|U^n(t, s)\mathcal{P}^n \varphi - U(t, s)\varphi\|_Z \\ & \leq \| \{ \mathcal{T}_0^n(t-s)\mathcal{P}^n \varphi - \mathcal{T}_0(t-s)\varphi \} \\ & \quad + \int_s^t \{ \mathcal{T}_0^n(t-r)\mathcal{B}^n(r)U^n(r, s)\mathcal{P}^n \varphi - \mathcal{T}_0(t-r)\mathcal{B}(r)U(r, s)\varphi \} dr \|_Z \\ & \leq \| \mathcal{T}_0^n(t-s)\mathcal{P}^n \varphi - \mathcal{T}_0(t-s)\varphi \|_Z \\ & \quad + \int_s^t \| \mathcal{T}_0^n(t-r)\mathcal{P}^n \mathcal{B}(r) \{ U^n(r, s)\mathcal{P}^n \varphi - U(r, s)\varphi \} \|_Z dr \\ & \quad + \int_s^t \| \{ \mathcal{T}_0^n(t-r)\mathcal{P}^n - \mathcal{T}_0(t-r) \} \mathcal{B}(r)U(r, s)\varphi \|_Z dr \\ & \leq \| \mathcal{T}_0^n(t-s)\mathcal{P}^n \varphi - \mathcal{T}_0(t-s)\varphi \|_Z \\ & \quad + \int_s^t \| \{ \mathcal{T}_0^n(t-r)\mathcal{P}^n - \mathcal{T}_0(t-r) \} \mathcal{B}(r)U(r, s)\varphi \|_Z dr \\ & \quad + \alpha \int_s^t \| U^n(r, s)\mathcal{P}^n \varphi - U(r, s)\varphi \|_Z dr. \end{aligned}$$

The convergence result given in (4.9), the strong continuity of the mappings  $\tau \mapsto \mathcal{B}(\tau)$  and  $(\sigma, \tau) \mapsto U(\tau, \sigma)$ ,  $\Delta_T$  compact, and an application of the Gronwall inequality then imply the convergence in (4.10) for  $\varphi \in \text{Dom}(\mathcal{A})$ . The uniform exponential bound on  $U$  and  $U^n$  for all  $n = 1, 2, \dots$ , together with the fact that  $\text{Dom}(\mathcal{A})$  is dense in  $Z$  then yield the desired result for all  $\varphi \in Z$ .  $\square$

## 5. An Example and Numerical Results

In this section we illustrate the application of our scheme to the estimation of the thermal conductivity in a one dimensional nonlinear (strictly speaking, quasi-linear) heat equation. We note that we do not attempt to verify  $A_0$ -admissibility or  $A_0$ -persistence of excitation for the example presented below. The  $A_0$ -admissibility is determined by the regularity of the plant. Assumptions, albeit rather restrictive

ones, can be made which would guarantee that the state of the plant is sufficiently regular to yield  $A_0$ -admissibility. As for  $A_0$ -persistence of excitation, the condition specified in Definition 3.3 is rather difficult, if not impossible to check. It is possible, however, to study this condition and its characterization in some very simple cases. Such a study provides valuable insight into tuning our adaptive estimator when it is applied in the context of more complex examples. In the case of the two-space theory detailed in [5] and [16], an investigation of the persistence of excitation condition was reported on in [6]. An analogous study for the one-space theory treated here is currently under way and when completed, will be reported on elsewhere.

We consider the identification of the thermal conductivity,  $\bar{q}$ , in the one dimensional quasi-linear heat equation

$$(5.1) \quad \frac{\partial \bar{x}}{\partial t}(t, \eta) - \frac{\partial}{\partial \eta} \left\{ \bar{q} \left( \left| \frac{\partial \bar{x}}{\partial \eta}(t, \eta) \right| \right) \frac{\partial \bar{x}}{\partial \eta}(t, \eta) \right\} = \bar{f}(t, \eta), \quad 0 < \eta < 1, \quad t > 0,$$

together with the Dirichlet boundary conditions

$$(5.2) \quad \bar{x}(t, 0) = 0 \quad \text{and} \quad \bar{x}(t, 1) = 0, \quad t > 0,$$

and the initial conditions

$$(5.3) \quad \bar{x}(0, \eta) = \bar{x}_0(\eta), \quad 0 \leq \eta \leq 1.$$

We assume that  $\bar{x}_0 \in L_2(0, 1)$  and  $\bar{f}(t, \cdot) \in L_2(0, 1)$ .

Let  $H = L_2(0, 1)$  be endowed with the standard inner product, and define the Hilbert space  $\hat{Q}$  as follows. Let

$$\hat{Q} = \{ \varphi : \varphi \in H_{loc}^1(\mathbb{R}^+) \text{ and } \varphi, D\varphi \in L_\infty(\mathbb{R}^+) \}.$$

Define the inner product,  $\langle \cdot, \cdot \rangle_Q$ , on  $\hat{Q}$  by

$$(5.4) \quad \langle \varphi, \psi \rangle_Q = \int_0^\infty \omega_0(\theta) \varphi(\theta) \psi(\theta) d\theta + \int_0^\infty \omega_1(\theta) D\varphi(\theta) D\psi(\theta) d\theta, \quad \varphi, \psi \in \hat{Q},$$

where  $\omega_0, \omega_1 \in L_1(\mathbb{R}^+)$  are positive weighting functions. Let  $\|\cdot\|_Q$  denote the norm induced by the inner product given in (5.4), and define the Hilbert space  $Q$  to be the completion of the inner product space  $\{\hat{Q}, \langle \cdot, \cdot \rangle_Q, \|\cdot\|_Q\}$ . Let  $Dom(A) = H^2(0, 1) \cap H_0^1(0, 1)$  and for  $q \in Q$  and  $\varphi \in Dom(A)$ , set  $A(q)\varphi = -D\{q(D\varphi)D\varphi\}$ .

For our estimator dynamics, we use the sum of a linear constant coefficient heat conduction operator and a decay term with Neumann (insulated) boundary conditions. That is, we define  $A_0 : Dom(A_0) \subset H \rightarrow H$  by

$$A_0\varphi = \{-\alpha_1 D^2 + \alpha_0\}\varphi, \quad \varphi \in Dom(A_0),$$

where  $\alpha_0, \alpha_1 > 0$ , with

$$Dom(A_0) = \{\varphi : \varphi \in H^2(0, 1), D\varphi(0) = D\varphi(1) = 0\}.$$

It follows that for  $\varphi \in \text{Dom}(A_0)$  we have

$$\langle A_0 \varphi, \varphi \rangle \geq \alpha_0 |\varphi|^2.$$

It is also easily argued that  $\lambda + A_0 : \text{Dom}(A_0) \subset H \rightarrow H$  is surjective for any  $\lambda > -\alpha_0$ .

We approximate using linear B-splines. For  $n = 1, 2, \dots$ , let  $\{\varphi_j^n\}_{j=0}^n$  be the standard linear B-splines on the interval  $[0, 1]$  defined with respect to the uniform mesh  $\{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$ . That is, for  $i = 0, 1, 2, \dots, n$

$$\varphi_i^n(x) = \begin{cases} 1 - |nx - i|, & x \in [\frac{i-1}{n}, \frac{i+1}{n}], \\ 0, & \text{elsewhere on } [0, 1]. \end{cases}$$

Set  $H^n = \text{span}\{\varphi_j^n\}_{j=0}^n$ . For each  $n = 1, 2, \dots$ , let  $P_H^n$  denote the orthogonal projection of  $H$  onto  $H^n$ . Standard approximation results for spline functions (see [15]) can be used to establish that the strong convergence to the identity of  $P_H^n$  required by Assumption (A3) is satisfied.

Let  $x_n(t) \in \mathbb{R}^{n+1}$  be the coordinate vector for  $x^n(t)$  with respect to the basis  $\{\varphi_j^n\}_{j=0}^n$ . That is,

$$x^n(t) = \sum_{j=0}^n x_n(t)_j \varphi_j^n.$$

Let  $M_n$  denote the Gram matrix corresponding to the basis  $\{\varphi_j^n\}_{j=0}^n$ . We have

$$M_n = [M_n]_{i,j} = \left[ \int_0^1 \varphi_i^n(x) \varphi_j^n(x) dx \right] = \frac{1}{6n} \begin{bmatrix} 2 & 1 & 0 & & & & & 0 \\ 1 & 4 & 1 & 0 & & & & \\ 0 & 1 & 4 & 1 & & & & \\ 0 & & & & & & & \\ & & & & & & & 0 \\ & & & & & & 1 & 4 & 1 & 0 \\ & & & & & & 0 & 1 & 4 & 1 \\ 0 & & & & & & & 0 & 1 & 2 \end{bmatrix}.$$

Also, let  $K_n$  be the  $(n+1) \times (n+1)$  matrix given by

$$K_n = [K_n]_{i,j} = \left[ \int_0^1 D\varphi_i^n(x) D\varphi_j^n(x) dx \right] \\ = n \begin{bmatrix} 1 & -1 & 0 & & & & & 0 \\ -1 & 2 & -1 & 0 & & & & \\ 0 & -1 & 2 & -1 & & & & \\ 0 & & & & & & & \\ & & & & & & & 0 \\ & & & & & & -1 & 2 & -1 & 0 \\ & & & & & & 0 & -1 & 2 & -1 \\ 0 & & & & & & & 0 & -1 & 1 \end{bmatrix}.$$



We approximate the operator  $A_0$ , by the operator  $A_0^n$  constructed using a standard Galerkin approach. For this type of approximation, it is easily argued that Assumptions (A4) and (A5) are satisfied. It is also easily established that the matrix representation for the operator  $A_0^n$  with respect to the basis  $\{\varphi_j^n\}_{j=0}^n$ ,  $A_{0n}$ , is given by

$$A_{0n} = M_n^{-1}(\alpha_1 K_n + \alpha_0 M_n) = \alpha_1 M_n^{-1} K_n + \alpha_0 I_n,$$

where  $I_n$  denotes the  $(n+1) \times (n+1)$  identity matrix.

We also use linear B-splines to discretize  $Q$ . For each  $m = 1, 2, \dots$ , and each  $r > 0$ , let  $\{\psi_j^{m,r}\}_{j=0}^m$  be the standard linear B-splines on the interval  $[0, r]$  defined with respect to the uniform mesh  $\{0, \frac{r}{m}, \frac{2r}{m}, \dots, r\}$ . Let  $Q^{m,r} = \text{span}\{\psi_j^{m,r}\}_{j=0}^m$ , where

$$\psi_j^{m,r} = \begin{cases} \hat{\psi}_j^{m,r}, & j = 0, 1, 2, \dots, m-1, \\ \hat{\psi}_m^{m,r} + \chi_{[r,\infty)}, & j = m, \end{cases}$$

with  $\chi_J$  denoting the characteristic function for the interval  $J$ . If we let  $P_Q^{m,r}$  denote the orthogonal projection of  $Q$  onto  $Q^{m,r}$ , the requisite strong convergence to the identity can be demonstrated. Consequently, Assumption (A3) holds. Let  $\Omega_{m,r}$  denote the  $(m+1) \times (m+1)$  Gram matrix corresponding to the basis  $\{\psi_j^{m,r}\}_{j=0}^m$ . That is

$$\begin{aligned} \Omega_{m,r} &= [\Omega_{m,r}]_{i,j} = \langle \psi_i^{m,r}, \psi_j^{m,r} \rangle_Q \\ &= \int_0^\infty \omega_0(\theta) \psi_i^{m,r}(\theta) \psi_j^{m,r}(\theta) d\theta + \int_0^\infty \omega_1(\theta) D\psi_i^{m,r}(\theta) D\psi_j^{m,r}(\theta) d\theta. \end{aligned}$$

Let  $q_{m,r}(t) = [q_{m,r}(t)_0, \dots, q_{m,r}(t)_m]^T \in \mathbb{R}^{m+1}$  denote the coordinate vector for the approximating estimate  $q^{m,r}(t)$  with respect to the basis  $\{\psi_j^{m,r}\}_{j=0}^m$ . That is

$$q^{m,r}(t) = \sum_{j=0}^m q_{m,r}(t)_j \psi_j^{m,r}, \quad t \geq 0.$$

There is some practical advantage to be gained by using a finite dimensional approximation to the plant data,  $\bar{x}(t)$ ,  $t \geq 0$ , in the estimator equations. The convergence theory given in Section 4 can be easily modified to permit this. (Our results on this will be reported on elsewhere.) To do this we use the orthogonal projection,  $P^n$ , of  $H_0^1(0,1)$  onto the subspace of  $H^n$ ,  $H_0^n$ , defined by  $H_0^n = \text{span}\{\varphi_j^n\}_{j=1}^{n-1}$ . If

$$\bar{x}^n(t) = P^n \bar{x}(t) = \sum_{j=1}^{n-1} \bar{x}_n(t)_j \varphi_j^n,$$

(i.e. let  $\bar{x}_n(t) \in \mathbb{R}^{n-1}$  be the coordinate vector for  $\bar{x}^n(t)$  with respect to the basis  $\{\varphi_j^n\}_{j=1}^{n-1}$ ) then

$$\bar{x}_n(t) = L_n^{-1} h_n(\bar{x}(t)),$$

where for  $\varphi \in H_0^1(0, 1)$ ,  $h_n(\varphi) \in \mathbb{R}^{n-1}$  is given by

$$h_n(\varphi)_i = \int_0^1 D\varphi(\eta) D\varphi_j''(\eta) d\eta, \quad j = 1, 2, \dots, n-1,$$

and  $L_n \in \mathbb{R}^{(n-1) \times (n-1)}$  is given by

$$L_n = [L_n]_{i,j} = \left[ \int_0^1 D\varphi_i''(x) D\varphi_j''(x) dx \right] = n \begin{bmatrix} 2 & -1 & 0 & & 0 \\ -1 & 2 & & & \\ 0 & & \ddots & & \\ & & & \ddots & 0 \\ 0 & & & 0 & -1 & 2 \end{bmatrix}.$$

It is easily verified that for  $\varphi \in H_0^1(0, 1)$

$$h_n(\varphi)_j = -n\Delta_n^2 \varphi\left(\frac{j-1}{n}\right) = -n \left\{ \varphi\left(\frac{j+1}{n}\right) - 2\varphi\left(\frac{j}{n}\right) + \varphi\left(\frac{j-1}{n}\right) \right\}, \\ j = 1, 2, \dots, n-1.$$

Thus the approximating estimator (i.e. (4.3)-(4.5) with  $\bar{x}$  replaced by  $\bar{x}^n$ ) does not require spatially distributed data. For a given value of  $n$ ,  $\bar{x}$  need only be spatially sampled at the  $n-1$  nodal points  $\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}$ .

For each  $t \geq 0$ , define the  $(n+1) \times (m+1)$  matrix  $B_{n,m,r}(t)$  by

$$[B_{n,m,r}(t)]_{i,j} = \int_0^1 \psi_j^{m,r}(D_\eta \bar{x}^n(t, \eta)) D_\eta \bar{x}^n(t, \eta) D\varphi_i''(\eta) d\eta, \\ j = 0, 1, 2, \dots, m, \quad i = 0, 1, 2, \dots, n.$$

Using the fact that  $\bar{x}^n(t) \in \text{span}\{\varphi_j^n\}_{j=1}^{n-1} \subset \text{span}\{\varphi_j^n\}_{j=0}^n$ , and the fact that  $D\varphi_j''$  is piecewise constant, and adopting the convention that  $\bar{x}_n(t)_0 = \bar{x}_n(t)_n = 0$ , we obtain that

$$[B_{n,m,r}(t)]_{i,j} = n\psi_j^{m,r}(n\{\bar{x}_n(t)_i - \bar{x}_n(t)_{i-1}\})\{\bar{x}_n(t)_i - \bar{x}_n(t)_{i-1}\} \\ - n\psi_j^{m,r}(n\{\bar{x}_n(t)_{i+1} - \bar{x}_n(t)_i\})\{\bar{x}_n(t)_{i+1} - \bar{x}_n(t)_i\}, \\ i = 1, 2, \dots, n-1, \quad j = 0, 1, 2, \dots, m,$$

and

$$[B_{n,m,r}(t)]_{i,j} = 0, \quad i = 0, n, \quad j = 0, 1, 2, \dots, m.$$

The matrix form of the approximating estimator, (4.3)-(4.5), is then given by

$$(5.5) \quad \dot{x}_n(t) + A_{0n}x_n(t) + M_n^{-1}B_{n,m,r}(t)q_{m,r}(t) \\ = M_n^{-1}\bar{f}_n(t) + A_{0n}\bar{x}_n(t), \quad t > 0,$$

$$(5.6) \quad \dot{q}_{m,r}(t) - \Omega_{m,r}^{-1}B_{n,m,r}(t)^T x_n(t) = -\Omega_{m,r}^{-1}B_{n,m,r}(t)^T \bar{x}_n(t), \quad t > 0,$$

$$(5.7) \quad x_n(0) = M_n^{-1}x_{0n}, \quad q_{m,r}(0) = \Omega_{m,r}^{-1}q_{0m,r},$$

where the  $(n+1)$ -vectors  $\bar{f}_n(t)$  and  $x_{0n}$  and the  $(m+1)$ -vector  $q_{0m,r}$  are given by

$$\bar{f}_n(t)_i = \int_0^1 \bar{f}(t, \eta) \varphi_i''(\eta) d\eta, \quad i = 0, 1, 2, \dots, n,$$

$$[x_{0n}]_i = \int_0^1 x_0(\eta) \varphi_i''(\eta) d\eta, \quad i = 0, 1, 2, \dots, n,$$

and

$$[q_{0m,r}]_i = \langle q_0, \psi_i^{m,r} \rangle_Q = \int_0^\infty \omega_0(\theta) q_0(\theta) \psi_i^{m,r}(\theta) d\theta + \int_0^\infty \omega_1(\theta) Dq_0(\theta) D\psi_i^{m,r}(\theta) d\theta, \quad i = 0, 1, \dots, m,$$

respectively.

We set

$$\bar{q}(\theta) = .9(1 - \frac{1}{2}e^{-\frac{1}{2}\theta^2}), \quad \theta \geq 0,$$

$$\bar{f}(t, \eta) = \{\sin(4\pi t) + 10^{-3}t^2\} \chi_{[.215, .315]}(\eta), \quad 0 < \eta < 1, \quad t > 0,$$

and

$$x_0(\eta) = 0, \quad 0 < \eta < 1,$$

and simulated the plant (equations (5.1)-(5.3)) using the IMSL routine DMOLCH, a double precision Hermite polynomial based method of lines PDE solver. In our estimator, we set  $\alpha_0 = 10^{-2}$ ,  $\alpha_1 = 5 \times 10^{-3}$ ,  $r = 3.5$ ,

$$\omega_0(\theta) = \omega_1(\theta) = \begin{cases} 1 & 0 \leq \theta < r \\ \frac{1}{2}e^{-20\theta} & r < \theta < \infty, \end{cases}$$

$$x_0(\eta) = 0, \quad 0 < \eta < 1,$$

and

$$q_0(\theta) = 1, \quad 0 < \theta < \infty.$$

We integrated the approximating estimator (i.e. equations (5.5)-(5.7)) using the NAG Library stiff system solver, D02NBF. The requisite integrals were computed using a composite two point Gauss-Legendre quadrature rule. All codes were written in FORTRAN and run on a Sun SPARCstation 10 under UNIX. In Figure 5.1 we have plotted our final (i.e. at time  $t = 100$ ) estimates for  $\bar{q}$  for various values of  $n$  and  $m$ . In Figure 5.2 we have plotted the estimates for  $\bar{q}$  at various times. These estimates were generated with  $n = 32$  and  $m = 16$ .

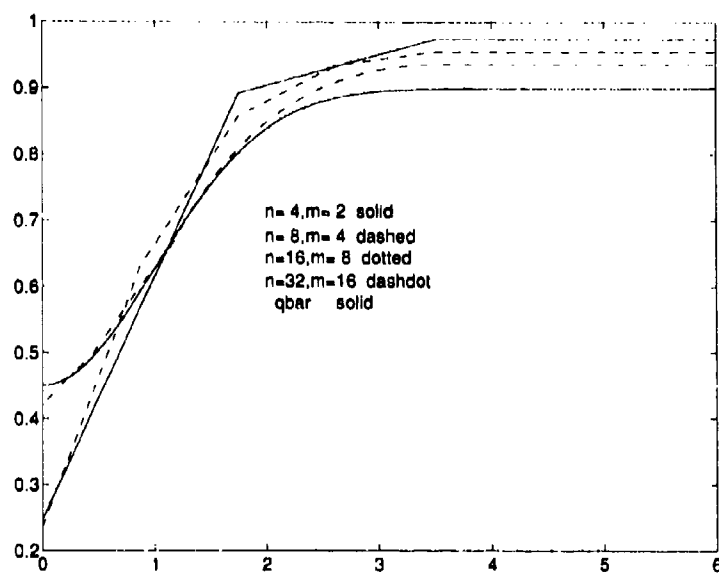


FIGURE 5.1. Final ( $t = 100$ ) estimates for  $\bar{q}$  for  $n = 2^{j+1}$ ,  $m = 2^j$  for  $j = 1, 2, 3, 4$ .

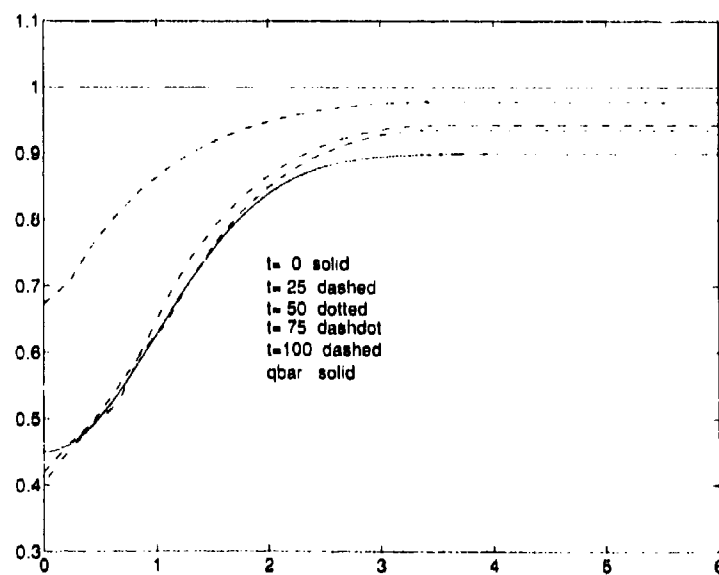


FIGURE 5.2. Estimates for  $\bar{q}$  at various times generated with  $n = 32$  and  $m = 16$ .

## References

1. H. W. Alt and K.-H. Hoffmann and J. Sprekels, *A numerical procedure to solve certain identification problems*, Intern. Series Numer. Math., 68 (1984), pp. 11-43.
2. J. Baumeister and W. Scordo, *Adaptive methods for parameter identification*, Methoden und Verfahren der Mathematischen Physik, (1987), pp. 87-116.
3. ———, *Asymptotic embedding methods for parameter estimation*, Proceedings of the 26th IEEE Conference on Decision and Control, (1987), pp. 170-174.
4. A. Belloni-Morante, *Applied Semigroups and Evolution Equations*, Clarendon Press, Oxford, 1979.
5. M. A. Demetrian and I. G. Rosen, *On-line parameter estimation for infinite dimensional dynamical systems*, CAMS Report 93-2, Center for Applied Mathematical Sciences, Department of Mathematics, University of Southern California, Los Angeles, CA 90089-1113, February 1993.
6. ———, *On the persistence of excitation in the adaptive identification of distributed parameter systems*, CAMS Report 93-4, Center for Applied Mathematical Sciences, Department of Mathematics, University of Southern California, Los Angeles, CA 90089-1113, March 1993 and *IEEE Transactions on Automatic Control*, to appear.
7. K.-H. Hoffmann and J. Sprekels, *The method of asymptotic regularization and restricted parameter identification problems in variational inequalities*, in Free Boundary Problems: Application and Theory, IV, Maubuisson, (1984), pp. 508-513.
8. ———, *On the identification of coefficients of elliptic problems by asymptotic regularization*, Numer. Funct. Anal. and Optimiz., 7 (1984-85), pp. 157-177.
9. ———, *On the identification of parameters in general variational inequalities by asymptotic regularization*, SIAM J. Math. Anal., 17 (1986), pp. 1198-1217.
10. T. Kato, *Perturbation Theory for Linear Operators*, Second Edition, Springer-Verlag, New York, 1984.
11. A. P. Morgan and K. S. Narendra, *On the stability of nonautonomous differential equations  $\dot{x} = [A + B(t)]x$ , with skew symmetric matrix  $B(t)$* , SIAM J. Control and Optimization 15 (1977), pp. 163-176.
12. K. S. Narendra and P. Kudva, *Stable adaptive schemes for system identification and control, Parts I and II*, IEEE Trans. Systems, Man and Cybernetics, SMC-4 (1974), pp. 542-560.
13. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer-Verlag, New York, 1983.
14. V. M. Popov, *Hyperstability of Control Systems*, Springer-Verlag, Berlin, Heidelberg, and New York, 1973.
15. M. H. Schultz, *Spline Analysis*, Prentice Hall, Englewood Cliffs, N.J., 1973.
16. W. Scordo, *Ein Modellgleichsverfahren zur adaptiven Parameteridentifikation in Evolutionsgleichungen*, PhD thesis, Johann Wolfgang Goethe-Universität zu Frankfurt am Main, Frankfurt am Main, Germany, 1987.

17. R. E. Showalter, *Hilbert Space Methods for Partial Differential Equations*, Pitman, London, 1977.
18. H. Tanabe, *Equations of Evolution*, Pitman, London, 1979.

Joseph Kazimir  
Center for Applied Mathematical Sciences  
Department of Mathematics  
University of Southern California  
Los Angeles, California 90089-1113, USA  
I.G. Rosen  
Center for Applied Mathematical Sciences  
Department of Mathematics  
University of Southern California  
Los Angeles, California 90089-1113, USA

## DECAY ESTIMATES FOR THE WAVE EQUATION WITH INTERNAL DAMPING

VILMOS KOMORNIK

Institut de Recherche Mathématique Avancée  
Université Louis Pasteur et C.N.R.S.

**ABSTRACT.** We obtain optimal estimates for the solution of an integral inequality related to many stabilization problems. Then it is applied to improve some recent results concerning the energy decay of the wave equation with internal nonlinear feedback. Unlike the earlier works, our method also applies in the case of bounded feedback functions.

1991 *Mathematics Subject Classification.* 35L05, 93D15

*Key words and phrases.* Wave equation, feedback stabilization, integral inequality.

### 1. Introduction

Let  $\Omega$  be a bounded domain in  $\mathbb{R}^n$  ( $n \geq 1$ ) having a boundary  $\Gamma$  of class  $C^\infty$  and let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a non-decreasing continuous function satisfying  $g(0) = 0$ . Consider the problem

$$(1.1) \quad u'' - \Delta u + g(u') = 0 \quad \text{in } \Omega \times \mathbb{R}_+$$

$$(1.2) \quad u = 0 \quad \text{on } \Gamma \times \mathbb{R}_+$$

$$(1.3) \quad u(0) = u_0 \quad \text{and} \quad u'(0) = u_1 \quad \text{on } \Omega$$

where we use the notation  $\mathbb{R}_+ := [0, +\infty)$ . We define the energy of the solutions of this system by

$$E = E(t) := \frac{1}{2} \int_{\Omega} u'(t)^2 + |\nabla u(t)|^2 \, dx, \quad t \in \mathbb{R}_+.$$

This problem is well-posed and dissipative. Indeed, we have the following theorem due to Haraux; it improved some earlier results of Lions–Strauss [11] and Brézis [1].

**Theorem 1.1 (Haraux [4], [5]).** *a) Given  $(u_0, u_1) \in H_0^1(\Omega) \times L^2(\Omega)$  arbitrarily, the problem (1.1)–(1.3) has a unique solution (defined in a suitable weak*

sense) satisfying

$$(1.4) \quad u \in C(\mathbb{R}_+; H_0^1(\Omega)) \cap C^1(\mathbb{R}_+; L^2(\Omega))$$

and

$$(1.5) \quad u'g(u') \in L_{loc}^1(\mathbb{R}_+; L^1(\Omega)).$$

Moreover, its energy is non-increasing.

b) If  $(u_0, u_1) \in (H^2(\Omega) \cap H_0^1(\Omega)) \times H_0^1(\Omega)$  and  $g(u_1) \in L^2(\Omega)$ , then the solution of (1.1)–(1.3) has the following regularity properties:

$$(1.6) \quad u \in L^\infty(\mathbb{R}_+; H^2(\Omega) \cap H_0^1(\Omega)),$$

$$(1.7) \quad u' \in L^\infty(\mathbb{R}_+; H_0^1(\Omega)),$$

$$(1.8) \quad u'' \in L^\infty(\mathbb{R}_+; L^2(\Omega)),$$

$$(1.9) \quad g(u') \in L^\infty(\mathbb{R}_+; L^2(\Omega)).$$

Moreover, the function

$$\mathbb{R}_+ \ni t \mapsto \int_{\Omega} |\Delta u(t, x)|^2 + |\nabla u'(t, x)|^2 dx$$

is non-increasing. Finally, the energy function is locally absolutely continuous and we have

$$(1.10) \quad E' = - \int_{\Omega} u'g(u') dx \leq 0$$

almost everywhere in  $\mathbb{R}_+$ .

The purpose of this paper is to study the energy decay rate under suitable growth assumptions on  $g$ . In the rest of the paper we assume that there exist three numbers  $p > 1$ ,  $q \geq 1 \geq r \geq 0$  and four positive constants  $c_i$ ,  $1 \leq i \leq 4$  such that

$$(1.11) \quad c_1|x|^p \leq |g(x)| \leq c_2|x|^{1/p} \quad \text{if } |x| \leq 1$$

and

$$(1.12) \quad c_3|x|^r \leq |g(x)| \leq c_4|x|^q \quad \text{if } |x| > 1.$$

Let us observe that if the inequalities of (1.11) (resp. of (1.12)) are satisfied only in some smaller neighbourhood of 0 (resp. of  $\pm\infty$ ), then they are also satisfied (maybe with other constants  $c_i$ ) in the neighbourhood  $|x| \leq 1$  (resp.  $|x| > 1$ ) by the non-decreasing property of  $g$ .



**Theorem 1.2.** Assume (1.11) and (1.12) with

$$(1.13) \quad p > 1, \quad r = 1 \quad \text{and} \quad q \geq 1 \quad \text{such that} \quad (n-2)q \leq n+2.$$

Then for every  $(u_0, u_1) \in H_0^1(\Omega) \cap L^2(\Omega)$  the solution of the problem (1.1) – (1.3) satisfies the estimate

$$(1.14) \quad E(t) \leq C t^{-2/(p-1)} \quad \forall t > 0,$$

where  $C$  is a constant depending only on the initial energy  $E(0)$  in a continuous way.

**Remarks.** --- There is an analogous result if we assume (1.11) with  $p = 1$ : then an easy adaptation of our proof given below leads to the estimate

$$(1.15) \quad E(t) \leq e^{1-\omega t} E(0) \quad \forall t > 0,$$

where  $\omega$  is a positive constant, independent of the initial data.

Theorem 1.2 improves and simplifies several earlier theorems of Nakao [13], Haraux and Zuazua [6], Zuazua [14], Conrad, Leblond and Marmorat [2] by weakening their growth assumptions on  $g$  near zero.

It is easy to give examples of functions  $g$  satisfying (1.11) – (1.13). For example, choose two numbers  $\alpha > 0$ ,  $\beta \geq 0$  and consider the function

$$(1.16) \quad g(x) := \begin{cases} |x|^\alpha \operatorname{sgn} x, & \text{if } |x| \leq 1, \\ |x|^\beta \operatorname{sgn} x, & \text{if } |x| > 1. \end{cases}$$

If  $n > 4$  and  $\beta > (n+4)/(n-4)$ , then we cannot apply theorem 1.1 because there is no  $q$  satisfying (1.12) and (1.13). We cannot apply this theorem either if  $\beta < 1$ : then the first inequality in (1.12) is not satisfied with  $r = 1$ . In the remaining cases, i.e. when  $\beta \geq 1$  and  $(n-4)\beta \leq n+4$ , we may apply theorem 1.1 with  $p = \max\{\alpha, 1/\alpha\}$ ,  $q = \beta$  and  $r = 1$ ; for  $p = \alpha = 1$  we rather apply the complement of the theorem mentioned above (cf.(1.15)).

As F. Conrad kindly pointed out to the author, theorem 1.2 and the previous results all have a drawback from the point of view of applications: they never apply if the function  $g$  is bounded. Indeed, no bounded function can verify the first inequality in (1.12) if  $r > 0$ . On the other hand, in many applications the function  $g$  is bounded, as for example the function introduced in (1.16) if  $\beta = 0$ . The following result applies in such situations as well, at least for smooth initial data.

**Theorem 1.3.** Assume (1.11), (1.12) with

$$(1.17) \quad p > 1, \quad 0 \leq r \leq 1 \quad \text{and} \quad q \geq 1 \quad \text{such that} \quad (n-4)q \leq n+4.$$

Furthermore, assume that

$$(1.18) \quad 2p \geq n + (2 - n)r \quad \text{and} \quad 2p \geq n - 2 - nq^{-1}.$$

Then for every  $(u_0, u_1) \in (H^2(\Omega) \cap H_0^1(\Omega)) \times H_0^1(\Omega)$  satisfying  $g(u_1) \in L^2(\Omega)$  there exists a constant  $C$  such that the solution of the problem (1.1)–(1.3) satisfies the estimate

$$(1.19) \quad E(t) \leq Ct^{-2/(p-1)} \quad \forall t > 0.$$

**Remarks.** — The technical condition (1.18) is easy to satisfy. Indeed, if (1.11), (1.12) and (1.17) are satisfied, then replacing  $p$  by

$$(1.20) \quad \max\{p, (n + (2 - n)r)/2, (n - 2 - nq^{-1})/2\}$$

(1.11), (1.12), (1.17) remain valid (with the same constants  $c_i$ ) and (1.18) will be satisfied, too. Note that condition (1.18) is automatically satisfied if  $n \leq 2$ .

Consider for example the function  $g$  defined by the formula (1.16) with  $\alpha > 0$ ,  $\beta \geq 0$ . If  $n > 4$  and  $\beta > (n + 4)/(n - 4)$ , then there is no  $q$  satisfying (1.12) and (1.17). On the other hand, if  $(n - 4)\beta \leq n + 4$ , then we may apply theorem 1.3 with  $r = \min\{\beta, 1\}$ ,  $q = \max\{\beta, 1\}$  and with any  $p > 1$  satisfying  $p \geq \max\{\alpha^{-1}/\alpha\}$  and (1.18). For example, if  $\alpha \neq 1$  and  $\beta = 0$ , then we may choose  $r = 0$ ,  $q = 1$  and  $p = \max\{\alpha, 1/\alpha, n/2\}$ .

Another example is given by

$$g(x) := \begin{cases} \sqrt{x}, & \text{if } x \geq 0, \\ -x^2, & \text{if } x \leq 0. \end{cases}$$

In this case theorem 1.2 cannot be applied because (1.12) is not satisfied with  $r = 1$ . If  $n > 12$ , then theorem 1.3 cannot be applied either because there is no  $q$  satisfying (1.12) and (1.17). On the other hand, in case  $n \leq 12$  theorem 1.3 applies with  $r = 1/2$ ,  $q = 2$  and with  $p = 2$  if  $n \leq 6$ ,  $p = (n + 2)/4$  if  $6 \leq n \leq 12$ . A more general example is given by the formula

$$g(x) := \sum_{i=1}^M c_i (x_+)^{p_i} - \sum_{i=1}^N d_i (x_-)^{q_i}$$

where  $x_+ = \max(x, 0)$ ,  $x_- = \max(-x, 0)$ ,  $c_i, d_i, p_i, q_i$  are given positive numbers.

There is a result analogous to theorem 1.3 if (1.11), (1.18) are satisfied with  $p = 1$  and  $q, r$  satisfy the conditions in (1.12), (1.17): if  $n \geq 2$ , then we also need the assumption  $r = 1$ . If these conditions are satisfied, then an easy adaptation of the proof of theorem 1.3 leads to the estimate (1.15) where  $\omega$  is a positive constant depending on the initial data.

Let us note that theorem 1.3 implies an earlier theorem of Nakao [13]; his original, stronger growth assumptions on  $g$  excluded the bounded functions.

Our method of proof differs from those of the earlier works. The idea is to prove that the energy function satisfies the integral inequalities

$$(1.21) \quad \int_t^\infty E^{\alpha+1}(s) \, ds \leq T E(0)^\alpha E(t), \quad \forall t \in \mathbb{R}_+$$

with  $\alpha = (p-1)/2$ , for a suitable positive constant  $A$ . Then we can conclude by applying the following result which improves some earlier theorems of Haraux [3] and Lagnese [10].

**Theorem 1.4.** — *Let  $E : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a non-increasing function and assume that there are two constants  $\alpha > 0$  and  $T > 0$  such that (1.21) is satisfied. Then we have*

$$(1.22) \quad E(t) \leq E(0) \frac{T + \alpha t^{-1/\alpha}}{T + \alpha T}, \quad \forall t \geq T.$$

**Remarks.** — Note that the inequality (1.22) is obviously satisfied for  $0 \leq t < T$  too; then it is weaker than the trivial inequality  $E(t) \leq E(0)$ . There is an analogous (and simpler) result if (1.21) is satisfied with  $\alpha = 0$ ; then we have

$$(1.23) \quad E(t) \leq E(0) e^{-t/T}, \quad \forall t \geq T,$$

instead of (1.22).

Our approach seems to be simpler than that of Nakao [13] because we apply simpler and more natural integral inequalities and also simpler than the Liapunov type method of Haraux and Zuazua [6] because we do not have to introduce artificial Liapunov functions and a corresponding artificial parameter.

Let us also mention that strong stability results (without decay estimates) were obtained earlier (by applying LaSalle's invariance principle) by Marcati [12].

The author is grateful to F. Conrad and A. Haraux for fruitful discussions and to the organizers of the 1993 Vöran Conference on Control Theory for their invitation.

## 2. Proof of Theorem 1.4

If  $E(0) = 0$ , then  $E \equiv 0$  and there is nothing to prove. Otherwise, replacing the function  $E$  by the function  $E/E(0)$  we may assume that  $E(0) = 1$  and we have to prove the estimate

$$(2.1) \quad E(t) \leq \frac{T + \alpha t^{-1/\alpha}}{T + \alpha T}, \quad \forall t \geq T.$$

Introduce the function

$$F: \mathbb{R}_+ \rightarrow \mathbb{R}_+, \quad F(t) = \int_t^\infty E^{\alpha+1} ds.$$

It is non-increasing and locally absolutely continuous. Differentiating and using (1.21) we find that

$$-F' \geq T^{-\alpha-1} F^{\alpha+1} \quad \text{a.e. in } (0, +\infty)$$

whence

$$(F^{-\alpha})' \geq \alpha T^{-\alpha-1} \quad \text{a.e. in } (0, B), \quad B := \sup\{t: E(t) > 0\}.$$

(Observe that  $F^{-\alpha}(t)$  is defined for  $t < B$ .) Integrating in  $[0, s]$  we obtain

$$(F(s))^{-\alpha} - F(0)^{-\alpha} \geq \alpha T^{-\alpha-1} s \quad \text{for every } s \in [0, B)$$

whence

$$(2.2) \quad F(s) \leq (F(0)^{-\alpha} + \alpha T^{-\alpha-1} s)^{-1/\alpha} \quad \text{for every } s \in [0, B).$$

Since  $F(s) = 0$  if  $s \geq B$ , this inequality is valid in fact for every  $s \in \mathbb{R}_+$ . Since  $F(0) \leq TE(0)^{\alpha+1} = T$  by (1.21), the right-hand side of (2.2) is less than equal to

$$(T^{-\alpha} + \alpha T^{-\alpha-1} s)^{-1/\alpha} = T^{(\alpha+1)/\alpha} (T + \alpha s)^{-1/\alpha}.$$

On the other hand,  $E$  being nonnegative and non-increasing, the left-hand side of (2.2) may be estimated as follows:

$$F(s) = \int_s^{+\infty} E^{\alpha+1} dt \geq \int_s^{T+(\alpha+1)s} E^{\alpha+1} dt \geq (T + \alpha s) E(T + (\alpha + 1)s)^{\alpha+1}.$$

Therefore we deduce from (2.2) the estimate

$$(T + \alpha s) E(T + (\alpha + 1)s)^{\alpha+1} \leq T^{(\alpha+1)/\alpha} (T + \alpha s)^{-1/\alpha}$$

whence

$$E(T + (\alpha + 1)s) \leq \left(1 + \frac{\alpha s}{T}\right)^{-1/\alpha}, \quad \forall s \geq 0.$$

Choosing  $t := T + (\alpha + 1)s$  hence the inequality (2.1) follows.  $\square$

**Remark 2.1.** — The theorem is optimal in the following sense: given  $\alpha > 0$ ,  $T > 0$ ,  $C > 0$  and  $t' \geq T$  arbitrarily, there exists a non-increasing function  $E: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  satisfying (1.21) and such that

$$E(0) = C \quad \text{and} \quad E(t') = E(0) \frac{T + \alpha t'^{-1/\alpha}}{T + \alpha T}.$$

We leave to the reader to verify that the following example has these properties:

$$(2.3) \quad E(t) := \begin{cases} C(1 + \alpha C^{-\alpha} t/T)^{-1/\alpha}, & \text{if } 0 \leq t \leq (t' - TC^\alpha)/(\alpha + 1), \\ C(1 + \alpha)^{1/\alpha} (1 + \alpha C^{-\alpha} t'/T)^{-1/\alpha}, & \text{if } (t' - TC^\alpha)/(\alpha + 1) \leq t \leq t', \\ 0, & \text{if } t > t'. \end{cases}$$

Let us also note that for  $t < T$  we cannot say more than the trivial estimate  $E(t) \leq E(0)$ . Indeed, for any given  $\alpha > 0$ ,  $T > 0$ ,  $C > 0$  and  $t' < T$  the function

$$(2.4) \quad E(t) := \begin{cases} C, & \text{if } 0 \leq t \leq T, \\ 0, & \text{if } t > T \end{cases}$$

satisfies (1.21) and  $E(t') = E(0) = C$ .

**Remark 2.2.** — Assume that  $E$  is also continuous. Then the inequalities (1.22) are strict; in particular,  $E(T) < E(0)$ . We refer to [9] for this result, for a detailed study of integral inequalities of type (1.21) (also for  $\alpha \leq 0$ ) and for the study of closely related differential inequalities.

### 3. Proof of Theorem 1.2

Applying a density and approximation argument as in [4], [5] it is sufficient to consider the case where

$$(u_0, u_1) \in (H^2(\Omega) \cap H_0^1(\Omega)) \times H_0^1(\Omega) \quad \text{and} \quad g(u_1) \in L^2(\Omega).$$

In this case the solution of (1.1) – (1.3) has the regularity properties (1.6) – (1.9) and this regularity is sufficient to justify all computations that follow. Let us begin by recalling the proof of formula (1.10) in theorem 1.1. Multiplying (1.1) with  $u'$ , integrating by parts in  $\Omega \times (S, T)$  and using (1.2) we obtain easily that

$$(3.1) \quad E(S) - E(T) = \int_S^T \int_{\Omega} u' g(u') \, dx \, dt, \quad 0 \leq S < T < +\infty.$$

Since  $xg(x) \geq 0$  for all  $x \in \mathbb{R}$ , it follows that the energy is non-increasing, locally absolutely continuous and

$$(3.2) \quad E' = - \int_{\Omega} u' g(u') \, dx \quad \text{a.e. in } \mathbb{R}_+.$$

Next we multiply (1.1) with  $E^{(p+1)/2}u$ . We have

$$\begin{aligned} 0 &= \int_S^T E^{(p+1)/2} \int_{\Omega} u(u'' - \Delta u + g(u')) \, dx \, dt \\ &= \left[ E^{(p+1)/2} \int_{\Omega} uu' \, dx \right]_S^T - \frac{p-1}{2} \int_S^T E^{(p-3)/2} E' \int_{\Omega} uu' \, dx \, dt \\ &\quad + \int_S^T E^{(p+1)/2} \int_{\Omega} |\nabla u|^2 - (u')^2 + ug(u') \, dx \, dt \end{aligned}$$

whence

$$\begin{aligned} \int_S^T E^{(p+1)/2} \, dt &= \left[ E^{(p+1)/2} \int_{\Omega} uu' \, dx \right]_S^T + \frac{p-1}{2} \int_S^T E^{(p-3)/2} E' \int_{\Omega} uu' \, dx \, dt \\ &\quad + \int_S^T E^{(p+1)/2} \int_{\Omega} 2(u')^2 - ug(u') \, dx \, dt \end{aligned}$$

Using the non-increasing property of  $E$ , the Sobolev imbedding  $H^1(\Omega) \subset L^2(\Omega)$ , the definition of  $E$  and the Cauchy-Schwarz inequality we have

$$\left| E^{(p+1)/2} \int_{\Omega} uu' \, dx \right| \leq c E^{(p+1)/2} \leq cE$$

and

$$\left| \int_S^T E^{(p-3)/2} E' \int_{\Omega} uu' \, dx \, dt \right| \leq -c \int_S^T E^{(p-1)/2} E' \, dt \leq -c \int_S^T E' \, dt \leq cE(S).$$

(Here and in the sequel we denote by  $c$  positive constants which may be different at different occurrences.) Using these estimates we conclude from the above identity that

$$(3.3) \quad 2 \int_S^T E^{(p+1)/2} \, dt \leq cE(S) + \int_S^T E^{(p-1)/2} \int_{\Omega} 2(u')^2 - ug(u') \, dx \, dt.$$

In order to estimate the last term of (3.3), we set

$$(3.4) \quad \Omega_1 := \{x \in \Omega : |u'| \leq 1\}, \quad \Omega_2 := \{x \in \Omega : |u'| > 1\},$$

and we remark that the assumption (1.11) implies the following two inequalities (we also use (3.2) and the Hölder inequality):

$$(3.5) \quad \int_{\Omega_1} (u')^2 \, dx \leq c \int_{\Omega_1} (u'g(u'))^{2/(p+1)} \, dx \leq c(-E')^{2/(p+1)},$$

and

$$(3.6) \quad \int_{\Omega_2} g(u')^2 \, dx \leq c \int_{\Omega_2} (u'g(u'))^{2/(p+1)} \, dx \leq c(-E')^{2/(p+1)},$$

Now fix an arbitrarily small  $\varepsilon > 0$  (to be chosen later). Using (3.5), (3.6), the boundedness of  $E^\alpha$  for any  $\alpha \geq 0$ , the obvious estimate

$$\|u\|_{L^2(\Omega)} \leq cE^{1/2}$$

and applying the Young inequality we obtain the following estimate:

$$\begin{aligned} E^{(p-1)/2} \int_{\Omega_1} 2(u')^2 - ug(u') \, dx \, dt &\leq cE^{(p-1)/2}(-E')^{2/(p+1)} + cE^{p/2}(-E')^{1/(p+1)} \\ &\leq \varepsilon E^{(p+1)/2} - c(\varepsilon)E'. \end{aligned}$$

Hence

$$(3.7) \quad \int_S^T E^{(p-1)/2} \int_{\Omega_1} 2(u')^2 - ug(u') \, dx \, dt \leq \varepsilon \int_S^T E^{(p+1)/2} \, dt + c(\varepsilon)E(S)$$

and substituting this into (3.3) we obtain that

$$(3.8) \quad (2-\varepsilon) \int_S^T E^{(p+1)/2} \, dt \leq c(\varepsilon)E(S) + \int_S^T E^{(p-1)/2} \int_{\Omega_2} 2(u')^2 - ug(u') \, dx \, dt.$$

Assume for a moment the estimates

$$(3.9) \quad E^{(p-1)/2} \int_{\Omega_2} (u')^2 \, dx \leq \varepsilon E^{(p+1)/2} - c(\varepsilon)E'$$

and

$$(3.10) \quad E^{(p-1)/2} \int_{\Omega_2} |ug(u')| \, dx \leq \varepsilon E^{(p+1)/2} - c(\varepsilon)E'.$$

Then we have (similarly to (3.7))

$$\int_S^T E^{(p-1)/2} \int_{\Omega_2} 2(u')^2 - ug(u') \, dx \, dt \leq 2\varepsilon \int_S^T E^{(p+1)/2} \, dt + c(\varepsilon)E(S)$$

and we conclude from (3.8) that

$$(3.11) \quad (2-3\varepsilon) \int_S^T E^{(p+1)/2} \, dt \leq c(\varepsilon)E(S).$$

Choosing  $\varepsilon = 1/3$  and  $A = c(\varepsilon)$  we deduce from (3.11) that

$$\int_S^T E^{(p+1)/2} \, dt \leq AE(S), \quad 0 \leq S \leq T < +\infty,$$

and letting  $T \rightarrow +\infty$  this yields the following crucial estimate

$$(3.12) \quad \int_S^\infty E^{(p+1)/2} \, dt \leq AE(S), \quad \forall S \geq 0.$$

We may thus complete the proof by applying theorem 1.4.

It remains to prove the estimates (3.9) and (3.10). Since now  $r = 1$  by hypothesis (3.9) follows easily from (1.12) and (3.2):

$$E^{(p-1)/2} \int_{\Omega_2} (u')^2 \, dx \leq c_3^{-1} \int_{\Omega_2} u' g(u') \, dx \leq -c_3^{-1} E'.$$

Turning to the proof of (3.10) first we apply the Hölder inequality, the Sobolev imbedding  $H^1(\Omega) \subset L^{q+1}(\Omega)$  (this follows from hypothesis (1.13) on  $q$ ) and the definition of the energy:

$$E^{(p-1)/2} \int_{\Omega_2} |u g(u')| \, dx \leq E^{(p-1)/2} \|u\|_{q+1} \|g(u')\|_{L^{(q+1)/q}(\Omega_2)} \leq c E^{p/2} \|g(u')\|_{L^{(q+1)/q}(\Omega_2)}.$$

(Here and in the sequel we denote the norm of  $L^\alpha(\Omega)$  by  $\|\cdot\|_\alpha$ ,  $1 \leq \alpha \leq \infty$ .) Using (1.12) and (3.2) to estimate the last norm we obtain that

$$E^{(p-1)/2} \int_{\Omega_2} |u g(u')| \, dx \leq c E^{p/2} \left( \int_{\Omega_2} u' g(u') \, dx \right)^{q/(q+1)} \leq c E^{p/2} |E'|^{q/(1+q)}.$$

Applying the Young inequality hence we conclude that

$$E^{(p-1)/2} \int_{\Omega_2} |u g(u')| \, dx \leq \varepsilon E^{p(q+1)/2} - c(\varepsilon) E'.$$

Since  $p(q+1) \geq p+1$ , hence (3.10) follows (with another  $\varepsilon$ ).

#### 4. Proof of Theorem 1.3

Adapting the proof of theorem 1.2 we only have to modify the proof of the estimates (3.9) and (3.10). Let us begin with the proof of (3.9). If  $r = 1$ , then the previous proof remains valid. If  $n = 1$ , then  $H^1(\Omega) \subset L^\infty(\Omega)$  and we conclude from (1.7) that  $u' \in L^\infty(\mathbb{R}_+; L^\infty(\Omega))$ . Using the obvious relation

$$|g(x)| \geq \min\{g(1), -g(-1)\} > 0 \quad \text{if } |x| > 1$$

it follows that

$$\begin{aligned} E^{(p-1)/2} \int_{\Omega_2} (u')^2 \, dx &\leq c E^{(p-1)/2} \int_{\Omega_2} |u'| |u' g(u')| \, dx \leq c E^{(p-1)/2} \|u'\|_\infty \|u' g(u')\|_1 \\ &\leq c E^{(p-1)/2} (-E') \leq -c E'. \end{aligned}$$

We may therefore assume that  $0 \leq r < 1$  and  $n \geq 2$ . First we show for every fixed  $\varepsilon' > 0$  and  $s \in (0, 1)$  the inequality

$$(4.1) \quad E^{(p-1)/2} \int_{\Omega_2} (u')^2 \, dx \leq \varepsilon' E^{(p-1)/(2(1-s))} \|u'\|_{(2-(r+1)s)/(1-s)}^{(2-(r+1)s)/(1-s)} - c(\varepsilon') E'.$$



Indeed, we have, putting  $\alpha := (2 - (p+1)s)/(1-s)$  for brevity,

$$\begin{aligned} E^{(p-1)/2} \int_{\Omega_2} (u')^2 \, dx &\leq c E^{(p-1)/2} \int_{\Omega_2} |u'|^{\alpha(1-s)} |u' g(u')|^s \, dx \\ &\leq c E^{(p-1)/2} \| |u'|^{\alpha(1-s)} \|_{1/(1-s)} \| (u' g(u'))^s \|_{1/s} \\ &= c E^{(p-1)/2} \| u' \|_{\alpha}^{\alpha(1-s)} \| u' g(u') \|_1^s \\ &= c E^{(p-1)/2} \| u' \|_{\alpha}^{\alpha(1-s)} (-E')^s \\ &\leq \varepsilon' E^{(p-1)/(2(1-s))} \| u' \|_{\alpha}^{\alpha} - c(\varepsilon') E'. \end{aligned}$$

Choosing  $s = 2/(p+1)$  the right hand side of (4.1) becomes

$$\varepsilon' E^{(p+1)/2} \| u' \|_{2(p-r)/(p-1)}^{2(p-r)/(p-1)} - c(\varepsilon') E'.$$

Assumption (1.18) on  $r$  implies the Sobolev imbedding  $H^1(\Omega) \subset L^{2(p-r)/(p-1)}(\Omega)$ . (This assumption is needed only if  $n \geq 3$ ; the case  $n = 2$  is trivial.) Using also (1.7) we thus deduce from (4.1) that

$$E^{(p-1)/2} \int_{\Omega_2} (u')^2 \, dx \leq c \varepsilon' E^{(p+1)/2} - c(\varepsilon') E'$$

with some constant  $c$  which does not depend on  $\varepsilon'$ . Choosing  $\varepsilon' := \varepsilon/c$  hence (3.9) follows.

Turning to the proof of (3.10) we may assume that  $q > 1$ . Indeed, if  $q = 1$ , then we may replace  $q$  by  $1 + \delta$  with a sufficiently small  $\delta > 0$ ; then (1.12), (1.17) and (1.18) remain valid. The case  $n < 4$  is obvious: then we have  $H^2(\Omega) \subset L^\infty(\Omega)$  whence  $u \in L^\infty(\mathbb{R}_+; L^\infty(\Omega))$  (cf. (1.6)). Therefore

$$\begin{aligned} E^{(p-1)/2} \int_{\Omega_2} |u g(u')| \, dx &\leq E^{(p-1)/2} \| u \|_{\infty} \| g(u') \|_{L^1(\Omega_2)} \\ &\leq c E^{(p-1)/2} \| u \|_{\infty} \| u' g(u') \|_{L^1(\Omega_2)} \\ &\leq c E^{(p-1)/2} \| u \|_{\infty} (-E') \leq -c E'. \end{aligned}$$

Now assume that  $n \geq 4$ . First we show for any fixed  $\varepsilon' > 0$  the inequality

$$(4.2) \quad E^{(p-1)/2} \int_{\Omega_2} |u g(u')| \, dx \leq \varepsilon' E^{(p-1)(1+q)/2} \| u \|_{q+1}^{q+1} - c(\varepsilon') E'.$$

Indeed, we have

$$\begin{aligned}
 E^{(p-1)/2} \int_{\Omega_2} |ug(u')| \, dx &\leq E^{(p-1)/2} \|u\|_{q+1} \|g(u')\|_{L^{1+1/q}(\Omega_2)} \\
 &\leq cE^{(p-1)/2} \|u\|_{q+1} \|(u'g(u'))^{q/(1+q)}\|_{L^{1+1/q}(\Omega_2)} \\
 &= cE^{(p-1)/2} \|u\|_{q+1} \|u'g(u')\|_{L^1(\Omega_2)}^{q/(1+q)} \\
 &\leq cE^{(p-1)/2} \|u\|_{q+1} (-E')^{q/(1+q)} \\
 &\leq \varepsilon' E^{(p-1)(1+q)/2} \|u\|_{q+1}^{q+1} - c(\varepsilon') E'.
 \end{aligned}$$

Next we recall that the interpolational inequality

$$(4.3) \quad \|u\|_{q+1} \leq c \|u\|_{2n/(n-2)}^t \|u\|_s^{1-t}$$

is valid for every  $s \in [2n/(n-2), +\infty)$  and  $t \in [0, 1]$  such that

$$(4.4) \quad \frac{1}{q+1} \geq t \frac{n-2}{2n} + \frac{1-t}{s}.$$

Choosing

$$t = \max \left\{ \frac{p+1}{q+1} + 1 - p, 0 \right\},$$

one can readily verify that  $t \in [0, 1]$ . Furthermore, it is easy to show that in case  $n = 4$  (4.4) is satisfied if we choose a sufficiently large  $s \geq 2n/(n-2)$ , while in case  $n > 4$  it follows from (1.17) and (1.18) that (4.4) is satisfied with  $s = 2n/(n-4)$ . Choosing  $s$  in this way, using (4.3), the Sobolev imbeddings  $H^1(\Omega) \subset L^{2n/(n-2)}(\Omega)$ ,  $H^2(\Omega) \subset L^s(\Omega)$  and applying (1.6) we obtain that

$$\begin{aligned}
 E^{(p-1)(1+q)/2} \|u\|_{q+1}^{q+1} &\leq cE^{(p-1)(1+q)/2} \|u\|_{2n/(n-2)}^{t(q+1)} \|u\|_{2n/(n-4)}^{(1-t)(q+1)} \\
 &\leq cE^{(p-1)(1+q)/2} \|u\|_{2n/(n-2)}^{t(q+1)} \leq cE^{(p-1+t)(1+q)/2} \\
 &\leq cE^{(p+1)/2}
 \end{aligned}$$

(in the last step we used the definition of  $t$ ). We thus deduce from (4.2) the estimate

$$E^{(p-1)/2} \int_{\Omega_2} |ug(u')| \, dx \leq c\varepsilon' E^{(p+1)/2} - c(\varepsilon') E'$$

with some constant  $c$  which does not depend on  $\varepsilon'$ . Choosing  $\varepsilon' := \varepsilon/c$  hence (3.10) follows.

Now the proof may be completed in the same way as that of theorem 1.2 in the preceding section.  $\square$

**Remarks.** — Unlike the situation of the preceding section, in the present case the constants in the estimates (3.9), (3.10) do not depend only on the initial energy of the solution. Therefore we cannot apply here a density argument to obtain the estimate (1.19) for all solutions with initial data in  $H_0^1(\Omega) \times L^2(\Omega)$ .

We mentioned in section 1 that there is a result, analogous to theorem 1.2 if  $p = 1$ . In this case we can repeat the above proof, excepting that we cannot assume in the case  $n \geq 2$  that  $r < 1$ . Under the additional assumption  $r = 1$  we may then prove (3.9) as follows: in the inequality following (4.1) we have  $\alpha = 2$  for any choice of  $s \in (0, 1)$ . Hence

$$\begin{aligned} E^{(p-1)/2} \int_{\Omega_2} (u')^2 \, dx &\leq \varepsilon' E^{(p-1)/(2(1-s))} \|u'\|_2^2 - c(\varepsilon') E' \\ &\leq c\varepsilon' E^{(p-1)/(2(1-s))} E - c(\varepsilon') E' \leq cE - c(\varepsilon') E' \end{aligned}$$

and (3.9) follows by choosing  $\varepsilon' := \varepsilon/c$ . Then we obtain again the integral estimates (3.12) (with  $p = 1$ ) and we conclude by applying the corresponding variant (with  $\alpha = 0$ ) of theorem 1.4.

## References

1. H. Brézis, *Problèmes unilatéraux*, J. Math. Pures Appl. **51** (1972), 1–168.
2. F. Conrad, J. Leblond and J. P. Marmorat, *Stabilization of second order evolution equations by unbounded nonlinear feedback*, to appear.
3. A. Haraux, *Oscillations forcées pour certains systèmes dissipatifs non linéaires*, preprint n° 78010, Laboratoire d'Analyse Numérique, Université Pierre et Marie Curie, Paris, 1978.
4. A. Haraux, *Comportement à l'infini pour une équation d'ondes non linéaire dissipative*, C. R. Acad. Sci. Paris Sér I Math. **287** (1978), 507–509.
5. A. Haraux, "Semilinear hyperbolic problems in bounded domains, Mathematical reports", J. Diendonné editor, Harwood Academic Publishers, Gordon and Breach, 1987.
6. A. Haraux and E. Zuazua, *Decay estimates for some semilinear damped hyperbolic problems*, Arch. Rat. Mech. Anal. (1988), 191–206.
7. V. Komornik, *Estimation d'énergie pour quelques systèmes dissipatifs semilinéaires*, C. R. Acad. Sci. Paris Sér I Math. **314** (1992), 747–751.
8. V. Komornik, *Stabilisation non linéaire de l'équation des ondes et d'un modèle de plaques*, C. R. Acad. Sci. Paris Sér I Math. **315** (1992), 55–60.
9. V. Komornik, *On some integral and differential inequalities*, preprint n°(1992) 190/P-283, IRMA, Université Louis Pasteur, Strasbourg, France.
10. J. Lagnese, *Boundary stabilization of thin plates*, SIAM Studies in Applied Mathematics **10**, SIAM, Philadelphia, PA, 1989.
11. J.-L. Lions and W. A. Strauss, *Some non-linear evolution equations*, Bull. Soc. Math. France **93** (1965), 43–96.
12. P. Marcati, *Decay and stability for nonlinear hyperbolic equations*, J. Diff. Equations **55** (1984), 30–58.
13. M. Nakao, *On the decay of solutions of some nonlinear dissipative wave equations in higher dimensions*, Math. Z. **193** (1986), 227–234.
14. E. Zuazua, *Stability and decay for a class of nonlinear hyperbolic problems*, Asymptotic Anal. **1**, 2 (1988), 161–185.

Vilmos Komornik  
Institut de Recherche Mathématique Avancée  
Université Louis Pasteur et C.N.R.S.  
7, rue René Descartes  
F-67084 Strasbourg Cedex, France

## ON THE CONTROLLABILITY OF THE ROTATION OF A FLEXIBLE ARM

W. KRABS

Fachbereich Mathematik  
Technische Hochschule Darmstadt

**ABSTRACT.** Considered is the rotation of a flexible arm in a horizontal plane about an axis through the arm's fixed end and driven by a motor whose torque is controlled. The model was derived and investigated computationally by Sakawa for the case that the arm is described as a homogeneous Euler beam. The resulting equation of motion is a partial differential equation of the type of a wave equation which is linear with respect to the state, if the control is fixed, and non-linear with respect to the control.

Considered is the problem of steering the beam, within a given time interval, from the position of rest for the angle zero into the position of rest under a certain given angle.

At first we show that, for every  $L^2$ -control which is suitably bounded, there is exactly one (weak) solution of the initial boundary value problem which describes the system without the end condition.

Then we present an iterative method for solving the problem of controllability and discuss its convergence.

1991 *Mathematics Subject Classification.* 93B05, 93C10, 93C15

*Key words and phrases.* Controllability, rotation of a flexible arm, homogeneous Euler beam.

### Introduction

The problem of controllability of a rotating homogeneous Euler beam has been dealt with in [4] and [5] where also references of former work are given. The model being investigated in [5] was slightly modified in [3] and also treated by a method of decomposition, however, of a more direct and simpler way than in [5].

In this paper we consider the model which was derived in [6] and treated numerically based on the method of Galerkin. We are concerned with the homogeneous case which leads to a differential equation of the form (1.8) with boundary conditions (1.3) and initial conditions (1.4) which correspond to the position of rest of the system at the beginning of the movement.

In Section 1 we formulate the problem of controllability to be solved in this paper. Section 2 is devoted to the question of solvability of the model equations (1.8), (1.3), (1.4). Here we also derive an integral differential equation (see (2.7))

which is equivalent to the model equations and show that this has a unique solution if the control is suitably bounded.

In Section 3 we present an iterative method for solving the problem of controllability and discuss its convergence under suitable conditions.

## 1. The Model and the Problem of Controllability

We consider the rotation of a flexible arm in a horizontal plane about an axis through the arm's fixed end. We assume this rotation to be driven by a motor whose torque is controlled. The equations of motion of the driving motor are given by

$$(1.1) \quad \begin{aligned} \dot{\varphi}(t) &= w(t), \\ \dot{w}(t) &= u(t), \quad t \in [0, T], \end{aligned}$$

for some given time  $T > 0$ , where  $\varphi = \varphi(t)$  is the angle of rotation,  $w = w(t)$  is the angular velocity, and  $u(t) = \tau(t)/J$  where  $\tau = \tau(t)$  is the torque generated by the motor and  $J$  is the moment of inertia of the motor. If the bending vibrations of the arm are small, we can assume  $J$  to be constant.

In addition we assume the arm to be homogeneous and of length 1. Following [6] we model the displacement  $y = y(t, x)$  of the arm from the rotating zero line by the differential equation

$$(1.2) \quad y_{tt}(t, x) + y_{xxxx}(t, x) - \frac{w(t)^2}{2} \frac{\partial}{\partial x} [(1 - x^2)y_x(t, x)] - w(t)^2 y(t, x) = -xu(t) \quad \text{for all } t \in (0, T), \quad x \in (0, 1).$$

The left end of the arm being clamped and the right end being free leads to the boundary conditions

$$(1.3) \quad \text{and} \quad \left. \begin{aligned} y(t, 0) &= y_x(t, 0) = 0 \\ y_{xx}(t, 1) &= y_{xxx}(t, 1) = 0 \end{aligned} \right\} \text{for } t \in [0, T].$$

At the beginning of the motion the arm is assumed to be in rest which leads to the initial conditions

$$(1.4) \quad y(0, x) = y_t(0, x) = 0 \quad \text{for } x \in (0, 1)$$

and

$$(1.5) \quad \varphi(0) = \dot{\varphi}(0) = 0.$$

On using (1.1) and (1.5) we get

$$(1.6) \quad \dot{\varphi}(t) = w(t) = \int_0^t u(s) ds$$

and

$$(1.7) \quad \varphi(t) = \int_0^t (t-s)u(s) ds$$

for  $t \in [0, T]$  so that (1.2) can be rewritten in the form

$$(1.8) \quad y_{tt}(t, x) + y_{xxxx}(t, x) - \left( \int_0^t u(s) ds \right)^2 \left\{ \frac{1}{2} \frac{\partial}{\partial x} [(1-x^2)y_x(t, x)] + y(t, x) \right\} \\ = -xu(t) \quad \text{for } t \in (0, T), \quad x \in (0, 1).$$

Now we are in the position of formulating the

**Problem of Controllability:** Let some angle  $\varphi_T \in \mathbb{R}$  with  $\varphi_T \neq 0$  be prescribed. Determine  $u \in L^2[0, T]$  such that

$$(1.9) \quad \int_0^T u(t) dt = 0 \quad (\Leftrightarrow \dot{\varphi}(T) = 0),$$

$$(1.10) \quad - \int_0^T tu(t) dt = \varphi_T \quad (\Leftrightarrow \varphi(T) = \varphi_T),$$

and the corresponding solution  $y = y(t, x)$  of (1.8), (1.3), and (1.4) satisfies the end condition

$$(1.11) \quad y(T, x) = y_t(T, x) = 0 \quad \text{for } x \in (0, 1).$$

In words: The torque of the driving motor is to be determined such that the arm is steered from a position of rest at  $t = 0$  with angle zero to a position of rest at  $t = T$  with angle  $\varphi_T$ .

## 2. On the Solvability of the Model Equations

Before dealing with the problem of controllability we investigate the question under which condition, for a given  $u \in L^2(0, T)$ , there exists a unique weak solution  $y = y(t, x)$  of (1.8), (1.3), (1.4).

For this purpose we at first observe that the differential operator

$$L[z](x) = z^{(4)}(x) \quad \text{for almost all } x \in (0, 1)$$

is self adjoint and positive definite on

$$D(L) = \{z \in H^4(0, 1) \mid z(0) = z'(0) = z''(1) = z'''(1) = 0\}$$

and has an infinite sequence of simple eigenvalues

$$\lambda_j = [(j - \frac{1}{2})\pi + \epsilon_j]^2, \quad j \in \mathbb{N},$$

which are the positive solutions of the transcendental equation

$$\cosh(\lambda_j^{\frac{1}{2}}) \cos(\lambda_j^{\frac{1}{2}}) = -1$$

and for which  $\lim_{j \rightarrow \infty} \epsilon_j = 0$  holds true. The corresponding sequence of orthonormal eigenfunctions  $e_j \in D(L)$ ,  $j \in \mathbb{N}$ , is given by  $(e_j = \varphi_j / \|\varphi_j\|_{L^2(0,1)})_{j \in \mathbb{N}}$  where

$$\varphi_j(x) = \cosh(\lambda_j^{\frac{1}{2}}x) - \cos(\lambda_j^{\frac{1}{2}}x) - \gamma_j(\sinh(\lambda_j^{\frac{1}{2}}x) - \sin(\lambda_j^{\frac{1}{2}}x))$$

and

$$\gamma_j = \frac{\cosh(\lambda_j^{\frac{1}{2}}) + \cos(\lambda_j^{\frac{1}{2}})}{\sinh(\lambda_j^{\frac{1}{2}}) + \sin(\lambda_j^{\frac{1}{2}})}, \quad j \in \mathbb{N}.$$

Now we consider the differential equation

$$(2.1) \quad \begin{aligned} y_{tt}(t, x) + y_{xxxx}(t, x) &= f(t, x) \\ \text{for } t &\in (0, T) \text{ and } x \in (0, 1) \end{aligned}$$

where  $f \in L^2((0, T), H)$  is chosen arbitrarily together with the boundary conditions (1.3) and the initial conditions (1.4). By results in [1] and [2] there is a unique weak solution  $y = y(t, x)$  of (2.1), (1.3), (1.4) with  $y \in C([0, T], V) \cap C^1([0, T], H)$  which is given by

$$(2.2) \quad y(t, x) = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} \int_0^t \int_0^1 f(s, \tilde{x}) e_j(\tilde{x}) d\tilde{x} \sin \sqrt{\lambda_j}(t-s) ds e_j(x)$$

for  $t \in [0, T]$  and  $x \in [0, 1]$  where  $H = L^2(0, T)$  and

$$V = \{v \in H^2(0, 1) \mid v(0) = v'(0) = 0\}.$$

Its time derivative reads

$$(2.3) \quad y_t(t, x) = \sum_{j=1}^{\infty} \int_0^t \int_0^1 f(s, \tilde{x}) e_j(\tilde{x}) d\tilde{x} \cos \sqrt{\lambda_j}(t-s) ds e_j(x)$$

for  $t \in [0, T]$  and  $x \in [0, 1]$ .

If  $y = y(t, x)$  is in  $C([0, T], V) \cap C^1([0, T], H)$  and is a solution of (1.8), (1.3), (1.4) for some given  $u \in L^2(0, T)$ , then we put

$$(2.4) \quad f(t, x) = \left( \int_0^t u(s) ds \right)^2 \left\{ \frac{1}{2} \frac{\partial}{\partial x} [(1-x^2)y_x(t, x)] + y(t, x) \right\} - xu(t)$$



for  $t \in (0, T)$  and  $x \in (0, 1)$  and infer that  $y = y(t, x)$  is also a weak solution of (2.1), (1.3), (1.4), since  $f \in L^2((0, T), H)$ .

In order to show that  $f \in L^2((0, T), H)$  we define

$$\begin{aligned} g(t, x) &= \frac{1}{2} \frac{\partial}{\partial x} [(1-x^2)y_x(t, x)] + y(t, x) \\ (2.5) \quad &= \frac{1}{2} (1-x^2)y_{xx}(t, x) - xy_x(t, x) + y(t, x) \end{aligned}$$

for  $t \in (0, T)$  and  $x \in (0, 1)$ . Then it follows, for every  $t \in (0, T)$ , that

$$\begin{aligned} \|g(t, \cdot)\|_H &\leq \frac{1}{2} \left( \int_0^1 (1-x^2)^2 y_{xx}(t, x)^2 dx \right)^{\frac{1}{2}} \\ (2.6) \quad &+ \left( \int_0^1 x^2 y_x(t, x)^2 dx \right)^{\frac{1}{2}} + \left( \int_0^1 y(t, x)^2 dx \right)^{\frac{1}{2}} \\ &\leq \frac{1}{2} \left( \int_0^1 y_{xx}(t, x)^2 dx \right)^{\frac{1}{2}} + 2 \left( \int_0^1 y_{xx}(t, x)^2 dx \right)^{\frac{1}{2}} + 4 \left( \int_0^1 y_{xx}(t, x)^2 dx \right)^{\frac{1}{2}} \\ &= \frac{13}{2} \|y(t, \cdot)\|_V. \end{aligned}$$

From this estimate and the fact that  $y \in C([0, T], V)$  one can easily infer that  $f$  defined by (2.4) is in  $L^2((0, T), H)$ . Therefore every weak solution  $y = y(t, x)$  of (1.8), (1.3), (1.4) for some given  $u \in L^2(0, T)$  which is in  $C([0, T], V) \cap C^1([0, T], H)$  satisfies the following integral differential equation

$$\begin{aligned} (2.7) \quad y(t, x) &= \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} \int_0^t \left( \int_0^s u(\sigma) d\sigma \right)^2 \int_0^1 g(s, \tilde{x}) c_j(\tilde{x}) d\tilde{x} \\ &\quad - \int_0^1 \tilde{x} c_j(\tilde{x}) d\tilde{x} u(s) \sin \sqrt{\lambda_j}(t-s) ds c_j(x) \end{aligned}$$

for  $t \in [0, T]$  and  $x \in [0, T]$  where  $g = g(t, x)$  is defined by (2.5).

Conversely, if  $y = y(t, x)$  is in  $C([0, T], V) \cap C^1([0, T], H)$  and satisfies (2.7) with  $g = g(t, x)$  being defined by (2.5), then  $y = y(t, x)$  is a weak solution of (1.8), (1.3), (1.4) in  $C([0, T], V) \cap C^1([0, T], H)$ .

For every  $u \in L^2(0, T)$  we define an affine linear mapping

$$S(u, \cdot) : C([0, T], V) \cap C^1([0, T], H) \rightarrow C([0, T], V) \cap C^1([0, T], H)$$

by

$$(2.8) \quad \begin{aligned} S(u, y)(t, x) = & \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} \int_0^t \left( \int_0^s u(\sigma) d\sigma \right)^2 \int_0^1 g(s, \tilde{x}, y) c_j(\tilde{x}) d\tilde{x} \\ & - u(s) \int_0^1 \tilde{x} c_j(\tilde{x}) d\tilde{x} \sin \sqrt{\lambda_j}(t-s) ds c_j(x) \end{aligned}$$

for  $(t, x) \in [0, T] \times [0, 1]$  where  $g = g(s, \tilde{x}, y)$  is defined by (2.5).

Obviously  $y \in C([0, T], V) \cap C^1([0, T], H)$  is a solution of (2.7), if and only if  $y$  satisfies the fixed point equation

$$(2.9) \quad y = S(u, y).$$

Let  $y_1, y_2 \in C([0, T], V) \cap C^1([0, T], H)$  be chosen arbitrarily. Then, for every  $t \in [0, T]$ , we obtain

$$\begin{aligned} & \|S(u, y_1)(t, \cdot) - S(u, y_2)(t, \cdot)\|_V^2 \\ &= \sum_{j=1}^{\infty} \left( \int_0^t \left( \int_0^s u(\sigma) d\sigma \right)^2 \int_0^1 g(s, \tilde{x}, y_1 - y_2) c_j(\tilde{x}) d\tilde{x} \sin \sqrt{\lambda_j}(t-s) ds \right)^2 \\ &\leq \sum_{j=1}^{\infty} \left( \int_0^t \left( \int_0^s u(\sigma) d\sigma \right)^4 \int_0^1 g(s, \tilde{x}, y_1 - y_2) c_j(\tilde{x}) d\tilde{x} \right)^2 ds \cdot T \\ &\leq T^4 \|u\|_H^4 \left( \operatorname{ess\,sup}_{t \in [0, T]} \|g(t, \cdot, y_1 - y_2)\|_H \right)^2 \\ &\leq \frac{169}{4} T^4 \|u\|_H^4 \left( \max_{t \in [0, 1]} \|(y_1 - y_2)(t, \cdot)\|_V \right)^2 \quad \text{see (2.6).} \end{aligned}$$

Therefore

$$(2.10) \quad \begin{aligned} & \max_{t \in [0, T]} \|S(u, y_1)(t, \cdot) - S(u, y_2)(t, \cdot)\|_V \\ &\leq \frac{13}{2} T^2 \|u\|_H^2 \max_{t \in [0, T]} \|(y_1 - y_2)(t, \cdot)\|_V \end{aligned}$$

which implies that equation (2.9) has exactly one solution  $y = y_u \in C([0, T], V) \cap C^1([0, T], H)$ , if there is some constant  $\gamma \in [0, 1]$  such that

$$\frac{13}{2} T^2 \|u\|_H^2 \leq \gamma^2$$

which is equivalent to

$$(2.11) \quad \|u\|_H \leq \sqrt{\frac{2}{13}} \frac{\gamma}{T}.$$

Under this condition  $y_u$  is also the unique weak solution of (1.8), (1.3), (1.4).

### 3. Iterative Solution of the Problem of Controllability

In view of the considerations of Section 2 the problem of controllability can be equivalently rewritten as follows: Find  $u \in L^2(0, T)$  and  $y \in Y = C([0, T], V) \cap C^1([0, T], H)$  such that

$$(3.1) \quad \int_0^T u(t) dt = 0, \quad \int_0^T t u(t) dt = -\varphi_T,$$

$$(3.2) \quad y - S(u, y) = 0,$$

$$(3.3) \quad y(T, \cdot) = y_t(T, \cdot) = 0$$

where

$$(3.4) \quad \begin{aligned} S(u, y)(t, x) = & \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} \int_0^t \left[ \left( \int_0^s u(\sigma) d\sigma \right)^2 \int_0^1 g(s, \tilde{x}, y) c_j(\tilde{x}) d\tilde{x} \right. \\ & \left. - u(s) \int_0^1 \tilde{x} c_j(\tilde{x}) d\tilde{x} \right] \sin \sqrt{\lambda_j}(t-s) ds c_j(x) \end{aligned}$$

for  $(t, x) \in [0, T] \times [0, 1]$  and

$$(3.5) \quad g(s, \tilde{x}, y) = \frac{1}{2} \frac{\partial}{\partial x} [(1 - \tilde{x}^2) g_x(s, \tilde{x})] + g(s, \tilde{x})$$

for  $(s, \tilde{x}) \in [0, T] \times [0, 1]$ .

We have shown that for each  $u \in H = L^2(0, T)$  with

$$(3.6) \quad \|u\|_H \leq \sqrt{\frac{2}{13}} \frac{\gamma}{T}$$

for some  $\gamma \in (0, 1)$  there is exactly one solution  $y = y_u \in Y$  of (3.2).

If  $u \in H$  with (3.1) and (3.6) is given, then the end conditions (3.3) for the corresponding solution  $y = y_u \in Y$  of (3.2) are equivalent to the following system of equations

$$\begin{aligned} \int_0^T \left[ \left( \int_0^t u(s) ds \right)^2 \int_0^1 g(t, \tilde{x}, y_u) c_j(\tilde{x}) d\tilde{x} - u(t) \int_0^1 \tilde{x} c_j(\tilde{x}) d\tilde{x} \right] \sin \sqrt{\lambda_j}(T-t) dt &= 0, \\ \int_0^T \left[ \left( \int_0^t u(s) ds \right)^2 \int_0^1 g(t, \tilde{x}, y_u) c_j(\tilde{x}) d\tilde{x} - u(t) \int_0^1 \tilde{x} c_j(\tilde{x}) d\tilde{x} \right] \cos \sqrt{\lambda_j}(T-t) dt &= 0 \end{aligned}$$

for  $j \in \mathbb{N}$ .

On using integration by parts in connection with

$$\int_0^T u(t) dt = 0$$

these turn out to be equivalent to

$$(3.7) \quad \left\{ \begin{array}{l} \int_0^T \left[ \int_0^t \int_0^s u(\sigma) d\sigma \int_0^1 g(s, \tilde{x}, y_u) c_j(\tilde{x}) d\tilde{x} \sin \sqrt{\lambda_j}(T-s) ds \right. \\ \quad \left. + \int_0^1 \tilde{x} c_j(\tilde{x}) d\tilde{x} \sin \sqrt{\lambda_j}(T-t) \right] u(t) dt = 0, \\ \int_0^T \left[ \int_0^t \int_0^s u(\sigma) d\sigma \int_0^1 g(s, \tilde{x}, y_u) c_j(\tilde{x}) d\tilde{x} \cos \sqrt{\lambda_j}(T-s) ds \right. \\ \quad \left. + \int_0^1 \tilde{x} c_j(\tilde{x}) d\tilde{x} \cos \sqrt{\lambda_j}(T-t) \right] u(t) dt = 0 \end{array} \right.$$

for  $j \in \mathbb{N}$ .

This leads to the following iterative method for solving the problem of controllability: At first we determine  $u_1 \in H$  such that

$$(3.8) \quad \int_0^T u_1(t) dt = 0, \quad \int_0^T t u_1(t) dt = -\varphi_T$$

and

$$(3.9) \quad \int_0^T \sin \sqrt{\lambda_j}(T-t) u_1(t) dt = \int_0^T \cos \sqrt{\lambda_j}(T-t) u_1(t) dt = 0.$$

This  $u_1$  is the first function in a sequence  $(u_k)_{k \in \mathbb{N}}$  in  $H$  which is constructed as follows: If  $u_k \in H$  has been determined, then  $y_k \in Y$  is determined as the unique solution of the equation

$$(3.10) \quad y_k - S(u_k, y_k) = 0.$$

Then  $u_{k+1} \in H$  is determined such that

$$(3.11) \quad \int_0^T u_{k+1}(t) dt = 0, \quad \int_0^T t u_{k+1}(t) dt = -\varphi_T.$$

$$(3.12) \quad \left\{ \begin{array}{l} \int_0^T \left[ \int_0^t \int_0^s u_k(\sigma) d\sigma \int_0^1 g(s, \tilde{x}, y_k) e_j(\tilde{x}) d\tilde{x} \sin \sqrt{\lambda_j}(T-s) ds \right. \\ \quad \left. + \int_0^1 \tilde{x} e_j(\tilde{x}) d\tilde{x} \sin \sqrt{\lambda_j}(T-t) \right] u_{k+1}(t) dt = 0, \\ \int_0^T \left[ \int_0^t \int_0^s u_k(\sigma) d\sigma \int_0^1 g(s, \tilde{x}, y_k) e_j(\tilde{x}) d\tilde{x} \cos \sqrt{\lambda_j}(T-s) ds \right. \\ \quad \left. + \int_0^1 \tilde{x} e_j(\tilde{x}) d\tilde{x} \cos \sqrt{\lambda_j}(T-t) \right] u_{k+1}(t) dt = 0 \end{array} \right.$$

for  $j \in \mathbb{N}$ .

Obviously we obtain (3.8) and (3.9) for  $k = 0$ , if we choose  $u_0 \equiv 0$  and  $y_0 \equiv 0$ .

Let us define, for every  $j \in \mathbb{N}$  and  $k \in \mathbb{N}_0$ ,

$$(3.13) \quad \left\{ \begin{array}{l} z_{jk}^1(t) = \int_0^t \int_0^s u_k(\sigma) d\sigma \int_0^1 g(s, \tilde{x}, y_k) e_j(\tilde{x}) d\tilde{x} \sin \sqrt{\lambda_j}(T-s) ds \\ \quad + \int_0^1 \tilde{x} e_j(\tilde{x}) d\tilde{x} \sin \sqrt{\lambda_j}(T-t), \\ z_{jk}^2(t) = \int_0^t \int_0^s u_k(\sigma) d\sigma \int_0^1 g(s, \tilde{x}, y_k) e_j(\tilde{x}) d\tilde{x} \cos \sqrt{\lambda_j}(T-s) ds \\ \quad + \int_0^1 \tilde{x} e_j(\tilde{x}) d\tilde{x} \cos \sqrt{\lambda_j}(T-t) \end{array} \right.$$

(with  $u_0 \equiv 0$  and  $y_0 \equiv 0$ ).

Then (3.12) can be written in the form

$$(3.14) \quad \int_0^T z_{jk}^1(t) u_{k+1}(t) dt = \int_0^T z_{jk}^2(t) u_{k+1}(t) dt = 0$$

for  $j \in \mathbb{N}$  and  $k \in \mathbb{N}$ .

Let  $W_k$  be the  $L^2$ -closure of the span of  $z_{0k} \equiv 1$ , and  $z_{jk}^1, z_{jk}^2$  for  $j \in \mathbb{N}$ .

**Assumption 1:** For every  $k \in \mathbb{N}$  the function  $f(t) = t$ ,  $t \in [0, T]$  does not belong to  $W_k$ .

For  $k = 0$  this follows from results in [2].

Under this Assumption there exists exactly one  $w_k \in W_k$  such that

$$\|w_k - f\|_{L^2(0,T)} \leq \|w - f\|_{L^2(0,T)} \quad \text{for all } w \in W_k$$

which is characterized by

$$\int_0^T (f(t) - w_k(t)) w(t) dt = 0 \quad \text{for all } w \in W$$

or equivalently by

$$\begin{aligned} \int_0^T (f(t) - w_k(t)) dt &= 0, \quad \int_0^T (f(t) - w_k(t)) z_{jk}^i(t) dt = 0 \\ &\text{for } i = 1, 2 \text{ and } j \in \mathbb{N}. \end{aligned}$$

Moreover, it follows that

$$\int_0^T (f(t) - w_k(t)) f(t) dt = \|f - w_k\|_{L^2(0,T)}^2 > 0.$$

If we put

$$(3.15) \quad u_{k+1} = \frac{\int_0^T (f(t) - w_k(t)) f(t) dt}{\|f - w_k\|_{L^2(0,T)}^2} (w_k - f),$$

then  $u_{k+1} \in H$  and satisfies (3.11) and (3.12).

**Assertion:** Among all solutions of (3.11) and (3.12) the one given by (3.15) has the smallest possible norm.

**Proof:** As a result of a well known duality theorem in linear approximation theory it follows that

$$\begin{aligned} \|f - w_k\|_{L^2(0,T)} &= \max \left\{ \left| \int_0^T f(t) u(t) dt \right| \mid u \in L^2(0,T), \int_0^T u(t) w(t) dt = 0 \right. \\ &\quad \left. \text{for all } w \in W_k \text{ and } \|u\|_{L^2(0,T)} = 1 \right\}. \end{aligned}$$

Let  $u \in L^2(0,T)$  be any solution of (3.11), (3.12). Then

$$\|u\|_{L^2(0,T)} > 0 \quad \text{and} \quad \int_0^T u(t) w(t) dt = 0 \quad \text{for all } w \in W_k.$$

Therefore

$$\left| \int_0^T f(t) \frac{u(t)}{\|u\|_{L^2(0,T)}} dt \right| = \frac{|\varphi_T|}{\|u\|_{L^2(0,T)}} \leq \|f - w_k\|_{L^2(0,T)} = \frac{|\varphi_T|}{\|u_{k+1}\|_{L^2(0,T)}}$$

which implies  $\|u_{k+1}\|_{L^2(0,T)} \leq \|u\|_{L^2(0,T)}$  and completes the proof.

**Assumption 2:** For every  $k \in \mathbb{N}_0$  we have

$$(3.16) \quad \|u_{k+1}\|_H \leq \sqrt{\frac{2}{13}} \frac{\gamma}{T}$$

for some  $\gamma \in (0, 1)$  where  $u_{k+1}$  is given by (3.15).

As an immediate consequence of this assumption we infer the existence of a subsequence  $(u_{k_i})_{i \in \mathbb{N}}$  of  $(u_k)_{k \in \mathbb{N}}$  which converges weakly to some  $\hat{u} \in H$ . Since the function  $u \mapsto \|u\|_{L^2(0,T)}$  is weakly sequentially lower semi-continuous, it follows that

$$\|\hat{u}\|_H \leq \liminf_{i \rightarrow \infty} \|u_{k_i}\|_H \leq \sqrt{\frac{2}{13}} \frac{\gamma}{T}.$$

Further it follows from (3.11) that

$$\int_0^T \hat{u}(t) dt = 0 \quad \text{and} \quad \int_0^T t \hat{u}(t) dt = -\varphi_T.$$

Let  $y_{\hat{u}} \in Y$  be the corresponding solution of (3.2). Then, for every  $k \in \mathbb{N}$  we get

$$y_{\hat{u}} - y_{u_k} = S(\hat{u}, y_{\hat{u}}) - S(u_k, y_{\hat{u}}) + S(u_k, y_{\hat{u}}) - S(u_k, y_{u_k})$$

which implies

$$\begin{aligned} \max_{t \in [0, T]} \|(y_{\hat{u}} - y_{u_k})(t, \cdot)\|_V &\leq \max_{t \in [0, T]} \| (S(\hat{u}, y_{\hat{u}}) - S(u_k, y_{\hat{u}}))(t, \cdot) \|_V \\ &\quad + \max_{t \in [0, T]} \| (S(u_k, y_{\hat{u}}) - S(u_k, y_{u_k}))(t, \cdot) \|_V \\ &\leq \max_{t \in [0, T]} \| (S(\hat{u}, y_{\hat{u}}) - S(u_k, y_{\hat{u}}))(t, \cdot) \|_V + \gamma^2 \max_{t \in [0, T]} \|(y_{\hat{u}} - y_{u_k})(t, \cdot)\|_V. \end{aligned}$$

Hence we obtain the estimate

$$(3.17) \quad \max_{t \in [0, T]} \|(y_{\hat{u}} - y_{u_k})(t, \cdot)\|_V \leq \frac{1}{1 - \gamma^2} \max_{t \in [0, T]} \| (S(\hat{u}, y_{\hat{u}}) - S(u_k, y_{\hat{u}}))(t, \cdot) \|_V.$$

Similar to the estimate (2.10) one can show that

$$\begin{aligned}
 (3.18) \quad & \max_{t \in [0, T]} \| (S(\hat{u}, y_{\hat{u}}) - S(u_k, y_{u_k}))(t, \cdot) \|_V \\
 & \leq \left( \frac{13}{2} T^2 \|\hat{u} + u_k\|_H \max_{t \in [0, T]} \|y_{u_k}(t, \cdot)\|_V + \sqrt{\frac{T}{3}} \right) \|\hat{u} - u_k\|_H \\
 & \leq (\sqrt{26} T \max_{t \in [0, T]} \|y_{u_k}(t, \cdot)\|_V + \sqrt{\frac{T}{3}}) \|\hat{u} - u_k\|_H.
 \end{aligned}$$

For every  $k \in \mathbb{N}$  we now define

$$\begin{aligned}
 \tilde{y}_k(t, x) = & \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} \int_0^t \int_0^s u_{k+1}(\sigma) d\sigma \int_0^s u_k(\sigma) d\sigma \int_0^1 g(s, \tilde{x}, y_k) c_j(\tilde{x}) d\tilde{x} \\
 & - \int_0^1 \tilde{x} c_j(\tilde{x}) d\tilde{x} u_{k+1}(s) \sin \sqrt{\lambda_j}(t-s) ds c_j(x)
 \end{aligned}$$

for  $(t, x) \in [0, T] \times [0, 1]$ .

Then it follows from (3.12) that

$$(3.19) \quad \tilde{y}_k(T, \cdot) = (\tilde{y}_k)_T(T, \cdot) = 0 \quad \text{a.e.}$$

for all  $k \in \mathbb{N}$ .

Further we obtain

$$(3.20) \quad \|(\tilde{y}_k - y_{u_k})(T, \cdot)\|_V \leq \frac{13}{2} T^2 \|u_k\|_H \max_{t \in [0, T]} \|y_{u_k}(t, \cdot)\|_V \|u_{k+1} - u_k\|_H$$

and

$$\|(\tilde{y}_k - y_{u_k})(T, \cdot)\|_H \leq \frac{13}{2} T^2 \|u_k\|_H \max_{t \in [0, T]} \|y_{u_k}(t, \cdot)\|_V \|u_{k+1} - u_k\|_H.$$

**Assumption 3:**  $\hat{u} = \lim_{k \rightarrow \infty} u_k$ . Then (3.17) and (3.18) imply

$$\lim_{k \rightarrow \infty} \max_{t \in [0, T]} \|(y_{\hat{u}} - y_{u_k})(t, \cdot)\|_V = 0$$

and on using (3.19) and (3.20) we infer

$$y_{\hat{u}}(T, \cdot) = (y_{\hat{u}})_T(T, \cdot) = 0 \quad \text{a.e.}$$

Hence  $\hat{u}$  and  $\hat{y}$  solve the problem of controllability.



## References

1. W. Krabs: On Time-Optimal Boundary Control of Vibrating Beams. In: Control Theory for Distributed Parameter Systems and Applications, edited by F. Kappel, K. Kunisch, W. Schappacher. Lecture Notes in Control and Information Sciences No 54, Springer-Verlag: Berlin Heidelberg New York Tokyo 1983, pp. 127 - 137.
2. W. Krabs: On Moment Theory and Controllability of One-Dimensional Vibrating Systems and Heating Processes. Lecture Notes in Control and Information Sciences No 173, Springer-Verlag: Berlin 1992.
3. W. Krabs: Controllability of a Rotating Beam. In: Analysis and Optimization of Systems: State and Frequency Domain Approaches for Infinite-Dimensional Systems, edited by R.F. Curtain. Lecture Notes in Control and Information Sciences No 185, Springer-Verlag: Berlin 1993, pp. 447 - 458.
4. G. Leugering: Control and Stabilization of a Flexible Robotarm. Dynamics and Stability of Systems 5 (1990), 37 - 46.
5. G. Leugering: On Control and Stabilization of a Rotating Beam by Applying Moments at the Base Only. In: Optimal Control of Partial Differential Equations, edited by K.-H. Hoffmann and W. Krabs. Lecture Notes in Control and Information Sciences No 449, Springer-Verlag: Berlin 1991, pp. 182 - 191.
6. Y. Sakawa, R. Ito and N. Fujii: Optimal Control of Rotation of a Flexible Arm. In: Control Theory for Distributed Parameter Systems and Applications, edited by F. Kappel, K. Kunisch, W. Schappacher. Lecture Notes in Control and Information Sciences No 54, Springer-Verlag: Berlin 1983, pp. 175 - 187.

W. Krabs  
Fachbereich Mathematik  
Technische Hochschule Darmstadt  
Schlossgartenstrasse 7  
D-64289 Darmstadt, West Germany

## MODELING AND CONTROLLABILITY OF INTERCONNECTED ELASTIC MEMBRANES

J. E. LAGNESE

Department of Mathematics  
Georgetown University

**ABSTRACT.** A collection of systems of partial differential equations, each of which models the nonlinear deformation of an elastic membrane, is considered. The state variables of the various membranes are required to satisfy certain "geometric" and "dynamic" coupling conditions which arise from continuity and balance law considerations. The resulting coupled systems then describe the nonlinear deformations of a system of interconnected membranes. The question of exact controllability of the linearization of such a network is discussed, where the controls are in the forms of forces applied on the outer edges and in the junction regions (where membranes are connected to one another).

1991 *Mathematics Subject Classification.* 35L20, 93B05, 93C20

*Key words and phrases.* Elastic membranes, controllability, junction conditions.

### 1. Modeling of Dynamic Nonlinear Elastic Membranes

Let  $\Omega$  be a bounded, open, connected set in  $\mathbb{R}^2$  with Lipschitz continuous boundary  $\Gamma$ , and let  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  be a fixed, right-handed orthonormal basis for  $\mathbb{R}^3$ . We denote by  $(x_1, x_2, x_3)$  (resp.,  $(x_1, x_2)$ ) the coordinates of a vector  $\mathbf{x} \in \mathbb{R}^3$  (resp.,  $\mathbf{x} \in \mathbb{R}^2$ ) with respect to the natural basis of  $\mathbb{R}^3$  (resp.,  $\mathbb{R}^2$ ). Consider an infinite elastic cylinder in  $\mathbb{R}^3$  whose reference configuration is

$$\mathcal{B} = \{\mathbf{p}_0 + x_i \mathbf{a}_i \mid (x_1, x_2) \in \Omega, -\infty < x_3 < \infty\},$$

where  $\mathbf{p}_0$  is a fixed vector in  $\mathbb{R}^3$ . Roman indices take the values 1, 2, 3. Greek indices take the values 1, 2, and summation convention is assumed. We set

$$\mathbf{r}(x_1, x_2, x_3) = \mathbf{p}_0 + x_i \mathbf{a}_i.$$

It is assumed that the motion of the cylinder obeys the following hypothesis.

---

Research supported by the Air Force Office of Sponsored Research through grant F49620-92-0031.

**Basic kinematic hypothesis.** The position vector at time  $t$  to the displaced location of the particle situated at  $\mathbf{r}(x_1, x_2, x_3)$  in the reference configuration is

$$(1.1) \quad \mathbf{R}(x_1, x_2, x_3, t) = \mathbf{r}(x_1, x_2, x_3) + \mathbf{W}(x_1, x_2, t).$$

Hypothesis (1.1) means that the deformation of a cross-section orthogonal to the axis of the cylinder is independent of the variable  $x_3$  along the axis of the cylinder. Thus the deformation of the cylinder is completely determined by the deformation of any of its cross-sections. The latter is specified by the vector function  $\mathbf{W}$  which measures the *displacement* of a cross-section. Let  $\{W_i\}$  denote the components of  $\mathbf{W}$  in the  $\{\mathbf{a}_i\}$  basis:  $\mathbf{W} = W_i \mathbf{a}_i$ . We refer to a cross-section as an *elastic membrane* and to the planar region

$$\mathcal{P} = \{\mathbf{p}_0 + x_\alpha \mathbf{a}_\alpha : (x_1, x_2) \in \Omega\}$$

as the *reference cross-section*.

Set  $\mathbf{G}_i = \mathbf{R}_{,i}$ , where a subscript after a comma means differentiation with respect to the indicated spatial variable. The components  $\varepsilon_{ij}$  of the *strain tensor* are given by

$$(1.2) \quad \varepsilon_{ij} = \frac{1}{2}(\mathbf{G}_i \cdot \mathbf{G}_j - \delta_{ij}).$$

It follows from (1.1) and (1.2) that

$$(1.3) \quad \begin{cases} \varepsilon_{\alpha\beta} = \frac{1}{2}(W_{\alpha,\beta} + W_{\beta,\alpha} + \mathbf{W}_{,\alpha} \cdot \mathbf{W}_{,\beta}), \\ \varepsilon_{\alpha 3} = \frac{1}{2}W_{3,\alpha}, \\ \varepsilon_{33} = 0. \end{cases}$$

**1.1. Equations of Motion.** Let  $s^{ij}$  denote the components of the *second Piola-Kirchhoff stress tensor*. This symmetric tensor relates the *stress vectors*  $\mathbf{t}^i$  to the quantities  $\mathbf{G}_i$  through

$$(1.4) \quad \mathbf{t}^i = s^{ij} \mathbf{G}_j$$

(see [8, Chapter 3]). If  $\mathbf{F}$  denotes the body force per unit of undeformed volume, the three-dimensional conditions for the balance of forces is

$$(1.5) \quad \mathbf{t}^i_{,i} + \mathbf{F} = \rho \mathbf{R}_{,tt},$$

where  $\rho$  is the mass of the body per unit of reference volume. Use of (1.1) and (1.4) in (1.5) allows that equation to be written

$$(1.6) \quad s^{ij}_{,j} \mathbf{a}_j + s^{i\beta}_{,\beta} \mathbf{W}_{,\beta} + s^{\alpha\beta} \mathbf{W}_{,\alpha\beta} + \mathbf{F} = \rho \mathbf{W}_{,tt}.$$

**Proposition 3.1.** *For any  $|\alpha| < 1$ , the corresponding eigenvalues of  $A_\alpha$  have the asymptotic form*

$$(3.1) \quad \lambda_k = \begin{cases} \frac{8\pi^3 |k|^3}{3\sqrt{3}} + O(k^2) & \text{if } \alpha = 0 \\ (12\pi^2 k^2 \ln|\alpha| + O(k)) + (8\pi^3 k^3 + O(k^2))i & \text{if } \alpha \neq 0, \end{cases}$$

as  $k \rightarrow \pm\infty$ . Moreover, there exists an  $\eta > 0$  such that

$$(3.2) \quad \operatorname{Re} \lambda_k > \eta, \quad \forall k.$$

**Proof:** Assume that  $\operatorname{Re} \lambda > 0$ . By symmetry we only need to consider the case that  $\operatorname{Im} \lambda \leq 0$ . We denote the three cube roots of  $-\lambda$  by  $\mu_0, \mu_1, \mu_2$ . These must have distinct real parts; we let  $\mu_0$  be the unique root such that  $0 < \arg(\mu_0) < \pi/6$  and

$$(3.3) \quad \mu_1 = e^{\frac{2\pi i}{3}} \mu_0 = \rho \mu_0, \quad \mu_2 = e^{\frac{4\pi i}{3}} \mu_0.$$

The general solution of the characteristic equation

$$(3.4) \quad \phi'''(x) + \lambda \phi(x) = 0$$

is then

$$(3.5) \quad \phi(x) = c_0 e^{\mu_0 x} + c_1 e^{\mu_1 x} + c_2 e^{\mu_2 x}$$

where  $c_0, c_1$  and  $c_2$  are arbitrary coefficients. Substituting (3.5) into the boundary conditions

$$(3.6) \quad \phi(1) = \phi(0) = 0, \quad \phi'(1) = \alpha \phi'(0)$$

we arrive at the system of equations

$$(3.7) \quad \begin{bmatrix} 1 & 1 & 1 \\ e^{\mu_0} & e^{\mu_1} & e^{\mu_2} \\ (\mu_0 - \alpha)\mu_0 & (\mu_1 - \alpha)\mu_1 & (\mu_2 - \alpha)\mu_2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = 0.$$

Setting the determinant of the matrix equal to zero we have

$$(3.8) \quad \begin{aligned} &(\mu_2 - \mu_1)e^{\mu_2 + \mu_1} + (\mu_0 - \mu_2)e^{\mu_0 + \mu_2} + (\mu_1 - \mu_0)e^{\mu_0 + \mu_1} + \alpha\mu_0(e^{\mu_0} - e^{\mu_1}) + \\ &\alpha\mu_1(e^{\mu_2} - e^{\mu_0}) + \alpha\mu_2(e^{\mu_0} - e^{\mu_1}) = 0. \end{aligned}$$

By our assumption,  $\operatorname{Re} \mu_0 > 0$ ,  $\operatorname{Re} \mu_1 < 0$ ,  $\operatorname{Re} \mu_2 < 0$ , and  $\operatorname{Re} \mu_0 \rightarrow +\infty$ ,  $\operatorname{Re} \mu_1 \rightarrow -\infty$  as  $|\lambda| \rightarrow \infty$ .

Multiplying both sides of (38) by  $e^{-\mu_1 u}$  and neglecting the terms  $\mu_1(\mu_1 u)^2 e^{-\mu_1 u}$ ,  $\mu_1^2 \mu_2 u e^{-\mu_1 u}$ ,  $\mu_1^2 \mu_2^2 u^2 e^{-\mu_1 u}$  and  $\mu_1^2 \mu_2^2 u^3 e^{-\mu_1 u}$  which are very small for large  $\lambda$ , we arrive at the equation

$$(\mu_1 - \mu_2)e^{-\mu_1 u} + \mu_1 - \mu_2 e^{-\mu_1 u} - \mu_1 \mu_2 e^{-\mu_1 u} = 0$$

which is, after simplification

$$(39) \quad 1 + \mu_1 u e^{-\mu_1 u} = \mu_2 e^{-\mu_1 u}.$$

If  $\mu_1 = 0$ , (39) reads  $1 = \mu_2 e^{-\mu_1 u}$ . Thus  $\mu_2 = 1/(1 + 2\mu_1 u)$  and

$$(40) \quad \lambda_1 = \lambda_2 = \lambda = \sum_{k=1}^{\infty} \frac{1}{k!} = 0.69,$$

as  $1 = \sum_{k=1}^{\infty} \frac{1}{k!}$ . If  $\lambda > 0$ , write  $\mu_1$  as  $\mu_1 = \lambda + \mu_2$ . Separating the real and imaginary parts of the equation (39), we obtain

$$(41) \quad \begin{cases} 1 - \cos(\mu_2 u) = (\lambda + \mu_2) \cos(\mu_2 u) e^{-\lambda u} - \mu_2 \cos(\mu_2 u) e^{-\lambda u} \\ \sin(\mu_2 u) = (\lambda + \mu_2) \sin(\mu_2 u) e^{-\lambda u} - \mu_2 \sin(\mu_2 u) e^{-\lambda u} \end{cases}$$

from which, by separating the real and imaginary parts of the equations and after dividing by

$$(42) \quad e^{-\lambda u}$$

$$\left( \frac{1}{2} + \frac{1}{2} \cos(\mu_2 u) \right) e^{-\lambda u} + \mu_2 \cos(\mu_2 u) e^{-\lambda u} = \left( \frac{1}{2} + \frac{1}{2} \cos(\mu_2 u) + \mu_2 \cos(\mu_2 u) \right) e^{-\lambda u}.$$

Note that  $\mu_2 = \lambda$  is the trivial solution of the problem  $\mu_1 = \lambda + \mu_2$  and  $\mu_2 = \lambda$ . There are no other solutions.

$$\cos(\mu_2 u) = \cos(\lambda u) \quad \text{and} \quad \sin(\mu_2 u) = \sin(\lambda u).$$

$$\mu_2 = \lambda + 2k\pi \quad \text{or} \quad \mu_2 = \lambda + (2k+1)\pi, \quad k \in \mathbb{Z}.$$

**Case (i):** when  $\mu_2 = \lambda + 2k\pi$ ,  $\lambda_1 = \lambda_2 = \lambda$  and

$$\lim_{\lambda \rightarrow \infty} \lim_{u \rightarrow \infty} \lambda_1 = \lim_{\lambda \rightarrow \infty} \lim_{u \rightarrow \infty} \lambda_2 = \lim_{\lambda \rightarrow \infty} \lim_{u \rightarrow \infty} \lambda = 0.$$

Thus since  $\lambda = \frac{\lambda_1}{2}$  and  $\cos \frac{\lambda_1}{2}$  that tend to infinity as

$$\lim_{\lambda \rightarrow \infty} \lim_{u \rightarrow \infty} \cos \frac{\lambda_1}{2} = \lim_{\lambda \rightarrow \infty} \lim_{u \rightarrow \infty} \cos \frac{\lambda}{2} = 1.$$

**Case (ii):** if  $\mu_2 = \lambda + (2k+1)\pi$ , the terms in (40) is smaller than other terms and we obtain approximately

$$e^{-\lambda u} = \mu_2 e^{-\lambda u}.$$

It yields that

$$\mu_2 = \lim_{\lambda \rightarrow \infty} \lim_{u \rightarrow \infty} \mu_2 = \lim_{\lambda \rightarrow \infty} \lim_{u \rightarrow \infty} \frac{1}{1 + 2k\pi} = 0.$$

**Case (iii):** letting  $\alpha \rightarrow +\infty$  in (3.12), we have that

$$\begin{aligned}\lim_{\alpha \rightarrow +\infty} (\cos(\alpha - \sqrt{\lambda/3} \sin(\alpha - \alpha/6)) &= 0 \\ \lim_{\alpha \rightarrow +\infty} (\sqrt{\lambda/3} \cos(\alpha - \alpha/6) - \sin(\alpha - \alpha/6)\sqrt{\lambda/3}) &= 0\end{aligned}$$

or equivalently

$$\lim_{\alpha \rightarrow +\infty} \sin(\alpha - \sqrt{\lambda/3} \alpha) = \lim_{\alpha \rightarrow +\infty} \cos(\alpha - \sqrt{\lambda/3} \alpha) = \frac{1}{2} \alpha$$

It leads to

$$\lim_{\alpha \rightarrow +\infty} \tan(\alpha - \sqrt{\lambda/3} \alpha) = \lim_{\alpha \rightarrow +\infty} \alpha = \infty$$

Thus

$$\mu_{k,\alpha} \rightarrow \ln(\alpha) + i\pi \frac{5\pi}{4} + 2k\pi + i\ln k$$

as  $\alpha \rightarrow +\infty$  and  $A_\alpha$  has eigenvalues

$$(3.13) \quad \lambda_k = \mu_{k,\alpha} - \mu_{-k,\alpha}^* = 8\pi^2 k^2 + O(k^2 \ln \alpha + 12\pi^2 k^2 \ln \alpha)$$

as  $\alpha \rightarrow +\infty$ . (3.10) and (3.13) give the formula (3.1) where the neglected terms referred to following (3.9) are corrected using the implicit function theorem: the one-to-one relationship between the eigenvalues and the indices  $k$  can be established using Rouche's theorem. An entirely similar argument gives the eigenvalues  $\lambda_{-k}$  as the conjugates of the  $\lambda_k$  shown.

To show (3.2), it suffices to prove that  $A_\alpha$  has no eigenvalue on the imaginary axis. Suppose that  $A_\alpha$  has an eigenvalue  $i\xi$  with  $\xi$  a real number and a corresponding eigenfunction  $g \in H_{0,1}^1$ , i.e.

$$(3.14) \quad A_\alpha g = i\xi g.$$

From the identity

$$(A_\alpha g, g)_{L^2(0,1)} = \frac{1}{2} (1 - \alpha^2) |g'(0)|^2,$$

one has  $g'(0) = 0$  and therefore  $g'(1) = 0$  since  $(A_\alpha g, g)_{L^2(0,1)} = i\xi \|g\|_{L^2(0,1)}^2$  and  $g'(1) = \alpha g'(0)$ . Write  $g$  as

$$g = g_1 + ig_2$$

with both  $g_1$  and  $g_2$  being real functions. Then it is from (3.14) that

$$(3.15) \quad g_1''' = \xi g_2, \quad g_2''' = -\xi g_1$$

with boundary conditions

$$(3.16) \quad g_k(1) = g_k(0) = g_k'(0) = g_k'(1) = 0, \quad k = 1, 2,$$

which force  $g = 0$ . Hence  $i\xi$  is not an eigenvalue of  $A_\alpha$ . The proof is complete.  $\square$

**Proposition 3.2.** Assume that  $\alpha \neq 0$ . Then the operator  $A_\alpha$  is a discrete spectral operator; all but a finite number of whose eigenvalues  $\lambda$  corresponds to one-dimensional projections  $E(\lambda, \lambda)$ . The operators  $A_\alpha, A'_\alpha$  with the indicated domains, have complete sets of eigenvectors, respectively,

$$\{e_k = (\lambda_k - \alpha)^{-1} k^{-1} e^{-i\lambda_k x}\}, \quad \{e_k = (\lambda_k - \alpha)^{-1} k^{-1} e^{i\lambda_k x}\}$$

which, normalized so that with  $\delta_{k,l}$ , the Kronecker delta,

$$(3.17) \quad (e_k, e_l)_{L^2(0,1)} = \delta_{k,l}$$

form dual Riesz-basis for  $L^2(0,1)$ .

**Proof:** The proof is almost same as those of Proposition 2.1 and 2.2 in [16] after the needed asymptotic form of the eigenvalues of  $A_\alpha$  has been established in Proposition 3.1.

**Remark 3.1.** The  $\alpha = 0$  is an exceptional case as that of  $\alpha = 1/2$  in [16]. It is not clear whether the corresponding eigenvectors form a Riesz-basis in the space  $L^2(0,1)$  in this exceptional case although its eigenvalues has faster growth rate.

**Definition 3.1.** We denote by  $H_\alpha^n$  the Hilbert space consisting of functions  $w$  in the Sobolev space  $H^n(0,1)$  which obey boundary conditions of the form

$$(3.18) \quad w^{(j)}(1) - w^{(j)}(0) = 0, \quad w^{(j-1)}(1) = \exp^{i\lambda_j} w^{(j-1)}(0),$$

as long as the indicated derivatives are of order  $j \leq n - 1$ . The norm and inner product in  $H_\alpha^n$  are  $\|\cdot\|_{H_\alpha^n}, (\cdot, \cdot)_{H_\alpha^n}$  inherited from  $H^n(0,1)$ .

The space  $H_\alpha^n$  can be characterized as follows using eigenvectors of  $A_\alpha$ .

**Proposition 3.3.** For  $\alpha \neq 0$ , a function  $w \in L^2$  also lies in  $H_\alpha^n$  if and only if it represented in the form

$$w = \sum_k c_k e_k$$

we have

$$(3.19) \quad \sum_k |k^n c_k|^2 < \infty.$$

Moreover,  $\|w\|_{H_\alpha^n}^2$  is a equivalent to the sum

$$\sum_k (|c_k|^2 + |k^n c_k|^2).$$

i.e., there are positive constants  $c_n, E_n$  (these will, in general, depend on  $\alpha$ ) such that, uniformly for  $w \in H_\alpha^n$ ,

$$(3.20) \quad c_n^2 \sum_{k=-\infty}^{\infty} [|c_k|^2 + |k^n c_k|^2] \leq \|w\|_{H_\alpha^n}^2 \leq E_n^2 \sum_{k=-\infty}^{\infty} (|c_k|^2 + |k^n c_k|^2).$$

**Proof:** The proof is similar to that of Proposition 2.4 in [16] and therefore is omitted here.

#### 4. Asymptotic Stability

In this section we consider asymptotic stability of solutions of the system (1.1) and show that small amplitude solutions decay exponentially in the space  $L^2(0, 1)$ . This is of course due to the dissipative mechanism introduced through the boundaries (cf. 1.3). Our approach is entirely same as the one used in [16] within the framework already established in the previous section.

We assume  $|\alpha| = 1$  and  $\alpha \neq 0$ . According the previous section, the resolution of the identity associated with the operator  $A_\alpha$  is the sum, strongly convergent in  $\mathcal{L}(L^2, L^2)$ ,

$$I = \sum_{k=-\infty}^{\infty} P_k$$

where  $P_k, -\infty < k < \infty$ , is the (generally non-orthogonal) projection defined by

$$P_k f = (f, c_k) c_{-k} c_k \quad \text{for any } f \in L^2(0, 1).$$

The corresponding strongly convergent representation of the semigroup generated by  $A_\alpha$  is

$$(4.1) \quad S_\alpha(t) = \sum_{k=-\infty}^{\infty} e^{N_k t} P_k,$$

where  $f_k = P_k f$ . From the representation (4.1) and the spectral properties of the operator  $A_\alpha$  it is easy to get the following proposition.

**Proposition 4.1.** *For any  $n > 0$ , there exists a  $B_n > 0$  and a  $\beta > 0$  such that*

$$(4.2) \quad \|S_\alpha(t)w_0\|_n \leq B_n e^{-\beta t} \|w_0\|_n$$

and

$$(4.3) \quad \int_0^\infty \|S_\alpha(t)w_0\|_{n+1}^2 dt \leq B_n^2 \|w_0\|_n^2$$



for any  $w_0 \in H_\alpha^n$  and  $t > 0$ . Besides,

$$(4.4) \quad \sup_{0 \leq t \leq \infty} \left\| \int_0^t S_\alpha(t-\tau)f(\cdot, \tau)d\tau \right\|_{n+1} \leq B_n \left( \int_0^\infty \|f(\cdot, \tau)\|_n^2 d\tau \right)^{\frac{1}{2}}$$

for any  $f \in L^2(0, \infty; H_\alpha^n)$ , and

$$(4.5) \quad \sup_{0 \leq t \leq \infty} \left\| \int_0^t S_\alpha(t-\tau)f(\cdot, \tau)d\tau \right\|_{n+1} \leq B_n \sup_{0 \leq t \leq \infty} \|f(\cdot, t)\|_n$$

for any  $f \in L^\infty(0, \infty; H_\alpha^n)$ .

Now we turn to consider the well-posedness of the nonlinear system (1.4). The global well-posedness of (1.4) in the space  $H_\alpha^n$  for  $n \geq 3$  has been established in section 2. However the approach there cannot be used to obtain the well-posedness of (1.4) in the space  $H_\alpha^n$  for  $n < 3$ . Here, based on the estimates established in Proposition 4.1, we are able to show that (1.4) is locally well-posed in the space  $H_\alpha^n$  for  $n = 1$  or 2.

**Theorem 4.1.** *Let  $n = 1$  or 2. Then for any  $\phi \in H_\alpha^n$  there exists a  $T = T(\|\phi\|_n) > 0$  such that the system (1.4) has a unique solution*

$$u \in X_T := C(0, T; H_\alpha^1) \cap L^\infty(0, T; H_\alpha^1)$$

where  $T \rightarrow \infty$  as  $\|\phi\|_1 \rightarrow 0$ . For any  $T' < T$ , there exists a neighborhood  $U$  of  $\phi$  in  $H_\alpha^n$  such that the map  $K: \phi \rightarrow u(\cdot, t)$  from  $U$  to  $X_{T'}$  is Lipschitz continuous.

**Proof:** It uses the same argument as that in [16].

Since the a priori global  $H^1$  or  $H^2$  estimates of solutions are not available, it is not clear whether (1.4) is global well-posed in the space  $H_\alpha^1$  or  $H_\alpha^2$ . However the small amplitude solutions of (1.4) do exist globally as shown by the following theorem.

**Theorem 4.2.** *Let  $n = 1$  or 2. There exists a  $\beta > 0$  such that for any  $\phi \in H_\alpha^n$  with  $\|\phi\|_n < \beta$ , (1.4) has a unique solution*

$$u \in C(0, \infty; H_\alpha^n) \cap L^\infty(0, \infty; H_\alpha^n)$$

**Proof:** see [16].

Then, using an infinite dimensional version of the second methods of Lyapounov as in [16] we may show that a small amplitude solution decays exponentially to 0.

**Theorem 4.3.** *There exists a  $\delta > 0$  such that for any  $\phi \in H_\alpha^1$  with  $\|\phi\|_1 < \delta$ , the corresponding unique solution of (1.4) satisfies*

$$(4.6) \quad \|u(\cdot, t)\|_{L^2(0,1)} \leq ce^{-\rho t} \|\phi\|_{L^2(0,1)}, \quad t > 0,$$

where  $c > 0$  and  $\rho > 0$  are independent of  $\phi$ .

**Proof:** see [16].

Finally we end this paper by presenting the following problem.

**Problem:** As we have known from Theorem 2.1 that the solution of (1.4) exists globally if its initial data is in the space  $H^3_\alpha$ . What is long time behavior of those solutions with large size of initial states?

The long time behavior of the solutions would reflect competition among the nonlinearity due to the term  $uu_x$ , dispersive effect due to the term  $u_{xxx}$  and the dissipative mechanism introduced through the boundary conditions. For the associated linear system, all solutions decay exponentially (see Proposition 4.1) which shows that the system possesses a rather strong dissipative mechanism. For the nonlinear system (1.4), Theorem 4.3 shows that a weak nonlinearity is not enough to overcome the introduced dissipative mechanism for the nonlinear term  $uu_x$  is even smaller comparing to the solution  $u$  if its initial data is small. However, for large amplitude solutions, the nonlinear term  $uu_x$  is expected to play a much more important role. It would be very interesting to see how a stronger nonlinearity to influence the long time behavior of the solutions of system (1.4).

## References

1. R. A. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.
2. J. L. Bona and R. Smith, The initial value problem for the Korteweg-de Vries equation, *Philos. Trans. Roy. Soc. London A*, **278** (1978), 555–601.
3. J. L. Bona and R. Winther, The Korteweg-de Vries equation in a quarter-plane, continuous dependence results, *Differential and Integral Equations* **2** (1989), 228–250.
4. B. A. Bubnov, Generalized boundary value problems for the Korteweg-de Vries equation in bounded domain, *Differential Equations* **15** (1979), 17–21.
5. B. A. Bubnov, Solvability in the large of nonlinear boundary value problem for the Korteweg-de Vries equations, *Differential Equations* **16** (1980), 21–30.
6. N. Dunford and J. T. Schwartz, *Linear Operators*, part III, Wiley-Interscience, 1971.
7. J. P. LaSalle and S. Lefschetz, *Stability by Lyapunov's direct method with applications*, Academic Press, New York, 1961.
8. T. Kato, On the Cauchy problem for the (generalized) Korteweg-de Vries equations, *Advances in Mathematics Supplementary Studies, Studies in Applied Math.* **8** (1983), 93–128.
9. C. E. Kenig, G. Ponce and L. Vega, Well-posedness of the initial value problem for the Korteweg-de Vries, *J. Amer. Math. Soc.* **4** (1991), 323–347.
10. ———, Well-posedness and scattering results for the generalized Korteweg-de Vries equation via the contraction principle, *Comm. Pure Appl. Math.* **46** (1993), 527–620.
11. V. Komornik, D. L. Russell and B.-Y. Zhang, Stabilisation de l'équation de Korteweg-de Vries, *C. R. Acad. Sci. Paris* **312** (1991), 841–843.
12. O. A. Ladyzenskaja, *Linear and Quasilinear Equations of Parabolic Type*, Translations of Mathematics Monograph, Vol. 23, AMS.
13. A. Pazy, *Semigroup of Linear Operators and Applications to Partial Differential Equations*, Applied Mathematical Sciences, Vol. 11, Springer-Verlag, 1983.
14. D. L. Russell, Nonharmonic Fourier series in the control theory of distributed parameter systems, *J. Math. Anal. Appl.* **18** (1967), 512–560.
15. D. L. Russell and B.-Y. Zhang, Controllability and stabilizability of the third order linear dispersion equation on a periodic domain, *SIMA Contr. & Opt.* **3** (1993), 659–676.

16. D. L. Russell and B.-Y. Zhang, Smoothing properties of solutions of the Korteweg-de Vries equation on a periodic domain with point dissipation, *IMA preprint*, series # 1085, December, 1992 (to appear in *J. Math. Anal. Appl.*).
17. D. L. Russell and B.-Y. Zhang, Stabilization of the Korteweg-de Vries equation on a periodic domain, *Control and Optimal Design of Distributed Parameter Systems*, edited by J. Lagnese, D. L. Russell and L. White, the IMA Volumes in Mathematics and Its Applications, Springer-Verlag, to appear.
18. J. C. Saut and R. Temam, Remarks on the Korteweg-de Vries equation, *Israel J. Math.* **24** (1976), 78–87.
19. B.-Y. Zhang, Taylor series expansion for solutions of the Korteweg-de Vries equation with respect to their initial values, *IMA preprint*, series # 1015, August 1992 (to appear in *J. Func. Anal.*).

Bing-Yu Zhang  
Department of Mathematical Sciences  
University of Cincinnati  
Cincinnati, Ohio 45221-0025, USA

## CONTROLLABILITY OF THE LINEAR SYSTEM OF THERMOELASTICITY: DIRICHLET-NEUMANN BOUNDARY CONDITIONS

ENRIQUE ZUAZUA

Departamento de Matemática Aplicada  
Universidad Complutense

**ABSTRACT.** We prove that the linear system of thermoelasticity with various boundary conditions is controllable in the following sense: If the control time is large enough and we act in the equations of displacement by means of a control supported in a neighborhood of the boundary of the thermoelastic body, then we may control exactly the displacement and simultaneously the temperature in an approximate way. We consider the following two cases: a) The displacement satisfies Dirichlet boundary conditions and the temperature takes Neumann zero boundary value; b) The displacement satisfies Neumann boundary conditions and the temperature vanishes at the boundary. The method of proof is inspired in our earlier work where the same result was proved for the case where both displacement and temperature satisfy Dirichlet boundary conditions.

1991 *Mathematics Subject Classification.* 93B05, 73C05, 35B37

*Key words and phrases.* Linear system of thermoelasticity, exact controllability, approximate controllability, decoupling, observability inequalities.

### 1. Introduction: First type of boundary conditions

Let us consider an isotropic and homogeneous thermoelastic body occupying an open and bounded set  $\Omega$  of  $\mathbb{R}^n$  ( $n \geq 1$ ) with boundary  $\Gamma := \partial\Omega$  of class  $C^2$ .

We denote by  $x = (x_1, \dots, x_n)$  a point of  $\Omega$  while  $t$  stands for the time variable. The displacement-vector is denoted by  $u = (u_1, \dots, u_n)$  ( $u_i = u_i(x, t)$ ,  $i = 1, \dots, n$ ) and the temperature by  $\theta = \theta(x, t)$ .

We fix a control time  $T > 0$  and a control region  $\omega$ : an open and non-empty subset of  $\Omega$ . We are allowed to act on the system through the equations of displacement by means of a control function

$$f = f(x, t) \in (L^2(\Omega \times (0, T)))^n$$

that represents an exterior force. The support of the control is restricted to the control region  $\omega$ . In the sequel, by  $\chi_\omega$  we denote the characteristic function of the set  $\omega$ .

The linear system of thermoelasticity with Dirichlet boundary conditions for the displacement and Neumann boundary conditions for the temperature and in the presence of this type of control  $f$  is as follows:

$$(1.1) \quad \begin{cases} u_{,t} - \mu \Delta u - (\lambda + \mu) \nabla \operatorname{div} u + \alpha \nabla \theta = f \chi_\omega & \text{in } \Omega \times (0, \infty) \\ \theta_t - \Delta \theta + \beta \operatorname{div} u_t = 0 & \text{in } \Omega \times (0, \infty) \\ u = 0, \quad \partial \theta / \partial \nu = 0 & \text{on } \Gamma \times (0, \infty) \\ u(x, 0) = u^0(x), u_t(x, 0) = u^1(x), \theta(x, 0) = \theta^0(x) & \text{in } \Omega \end{cases}$$

where  $\mu, \lambda > 0$  are Lamé's constants,  $\alpha, \beta \in \mathbb{R}$  ( $\alpha/\beta > 0$ ) the coupling parameters and  $\nu$  the outward unit normal to  $\Omega$  in  $\Gamma$ .

When  $f = 0$  system (1.1) is well-posed in

$$H = (H_0^1(\Omega))^n \times (L^2(\Omega))^n \times L^2(\Omega).$$

More precisely, for every initial data  $(u^0, u^1, \theta^0) \in H$  there exists a unique solution

$$(u, u_t, \theta) \in C([0, \infty); H).$$

This solution is given by

$$(u(t), u_t(t), \theta(t)) = S(t)(u^0, u^1, \theta^0), t > 0$$

where  $S(t) : H \rightarrow H, t > 0$ , is the strongly continuous semigroup generated by system (1.1). We will denote by  $S_i(t), i = 1, 2, 3$  the three components of  $S(t)$ .

On the other hand, the energy

$$(1.2) \quad E(t) = \frac{1}{2} \int_{\Omega} \left[ |u_t|^2 + \mu |\nabla u|^2 + (\lambda + \mu) |\operatorname{div} u|^2 + \frac{\alpha}{\beta} \theta^2 \right] dx$$

satisfies

$$\frac{dE(t)}{dt} = -\frac{\alpha}{\beta} \int_{\Omega} |\nabla \theta|^2 dx.$$

If  $f \in L^1(0, T; (L^2(\Omega))^n)$  system (1.1) possesses an unique solution  $(u, u_t, \theta)$  in  $C([0, T]; H)$  given in terms of the semigroup  $S$  by the variation of constants formula.

We consider the following controllability problem: Given  $(u^0, u^1, \theta^0)$  and  $(v^0, v^1, \eta^0)$  in  $H$  and  $\varepsilon > 0$ , to find a control  $f$  such that the solution of (1.1) satisfies

$$(1.3) \quad \begin{cases} u(T) = v^0, u_t(T) = v^1 \\ \|\theta(T) - \eta^0\|_{L^2(\Omega)} \leq \varepsilon. \end{cases}$$

In other words, we request the exact controllability of the displacement and the approximate controllability of the temperature.

In [Z3] we gave a positive answer to this question in the case where both displacement and temperature satisfy Dirichlet boundary conditions. However, in

the present case, the approximate controllability does not hold in any time  $T$  and whatever the support of the controls  $\omega$  is. This is due to the fact that the quantity

$$(1.4) \quad F(t) = \int_{\Omega} \theta(x, t) dx$$

remains constant in time along solution trajectories of (1.1). This can be checked easily integrating the second equation of (1.1) in  $\Omega$ .

The fact that  $F(\cdot)$  is preserved along trajectories suggests the following controllability problem: Given  $(u^0, u^1, \theta^0), (v^0, v^1, \eta^0) \in H$  such that

$$(1.5) \quad \int_{\Omega} \theta^0(x) dx = \int_{\Omega} \eta^0(x) dx$$

and  $\varepsilon > 0$ , to find a control  $f$  such that the solution of (1.1) satisfies (1.3).

In other words, we request the exact controllability of the displacement and the approximate controllability of the temperature at a prescribed level of total amount of heat.

In the sequel, if this property holds we will say that system (1.1) is exact-approximately controllable.

We have the following result.

**Theorem 1.1.** *Let  $\omega$  be a neighborhood of the boundary  $\Gamma$  in  $\Omega$ , i. e.  $\omega = \Omega \cap \Theta$  where  $\Theta$  is a neighborhood of  $\Gamma$  in  $\mathbb{R}^n$ . Suppose that  $T > \text{diam}(\Omega \setminus \omega)/\sqrt{\mu}$ . Then, system (1.1) is exact-approximately controllable in time  $T$ .*

**Remark 1.1.** Theorem 1.1 is a variant of the result proved in [Z3] in the case where both displacement and temperature satisfy Dirichlet boundary conditions. For the proof of Theorem 1.1 we will use basically the method developed in [Z3] (see also [Z1] and [Z2]). However, considering that we have different boundary conditions and that the constraint (1.5) is now present, some additional developments will be necessary. ■

**Remark 1.2.** In one space dimension the same result can be proved for any open subset  $\omega$  of  $\Omega$ . We refer to [Z3] for a detailed proof in the case where both displacement and temperature satisfy Dirichlet boundary conditions. ■

A natural way of addressing the presence of the constraint (1.5) is to decompose the temperature  $\theta$  so that

$$(1.6) \quad \theta(x, t) = \frac{1}{|\Omega|} \int_{\Omega} \theta^0(x) dx + \tilde{\theta}(x, t) dx$$

where  $\tilde{\theta}$  is such that  $(u, \tilde{\theta})$  satisfy (1.1) with the following modified initial data for  $\tilde{\theta}$ :

$$(1.7) \quad \tilde{\theta}(x, 0) = \tilde{\theta}^0(x) = \theta^0(x) - \frac{1}{|\Omega|} \int_{\Omega} \theta^0(x) dx.$$

We then introduce the Hilbert space  $\mathcal{H} = \{f \in L^2(\Omega) : \int_{\Omega} f(x) dx = 0\}$  and the new phase space for the system of thermoelasticity:  $\mathcal{V} = (H_0^1(\Omega))^n \times (L^2(\Omega))^n \times \mathcal{H}$ .

Since the space  $\mathcal{V}$  is invariant under the action of the semigroup  $S(t)$  even if an external force is added in the first equation of (1.1) (i. e. when  $f \neq 0$ ), then our controllability problem can be formulated in the following way: Given  $(u^0, u^1, \theta^0), (v^0, v^1, \eta^0) \in \mathcal{V}$  and  $\varepsilon > 0$ , to find a control  $f$  such that the solution of (1.1) satisfies (1.3).

One of the main ingredients of the proof of Theorem 1 is the following observability inequality for the adjoint system of thermoelasticity:

$$(1.8) \quad \begin{cases} \varphi_{tt} - \mu \Delta \varphi - (\lambda + \mu) \nabla \operatorname{div} \varphi + \beta \nabla \psi_t = 0 & \text{in } \Omega \times (0, T) \\ -\psi_t - \Delta \psi - \alpha \operatorname{div} \varphi = 0 & \text{in } \Omega \times (0, T) \\ \varphi = 0, \quad \partial \psi / \partial \nu = 0 & \text{on } \partial \Omega \times (0, T) \\ \varphi(x, T) = \varphi^0(x), \varphi_t(x, T) = \varphi^1(x) & \text{in } \Omega \\ \psi(x, T) = \psi^0(x) & \text{in } \Omega \end{cases}$$

where  $\varphi = (\varphi_1, \dots, \varphi_n)$  is the adjoint displacement variable and  $\psi$  the temperature.

**Proposition 1.1.** *Under the assumptions of Theorem 1.1, for every bounded set  $B$  of  $\mathcal{H}$  there exists  $\delta = \delta(B) > 0$  such that*

$$(1.9) \quad \delta \leq \int_0^T \int_{\Omega} |\varphi|^2 dx dt$$

holds for every solution of (1.8) with initial data such that

$$\|(\varphi^0, \varphi^1 + \beta \nabla \psi^0)\|_{(L^2(\Omega))^n \times (H^{-1}(\Omega))^n} \geq 1, \psi^0 \in B.$$

By  $P : (L^2(\Omega))^n \rightarrow (L^2(\Omega))^n$  we denote the continuous and linear operator defined as  $Pf = \nabla g$  where  $g \in H^1(\Omega) \cap \mathcal{H}$  is the unique solution of the following variational problem:

$$(1.10) \quad \int_{\Omega} \nabla g \cdot \nabla \rho dx = \int_{\Omega} f \cdot \nabla \rho dx, \quad \text{for all } \rho \in H^1(\Omega) \cap \mathcal{H}.$$

Obviously, the solution  $g$  of (1.10) in  $H^1(\Omega) \cap \mathcal{H}$  exists and it is unique. Note that when  $f \in (H_0^1(\Omega))^n$ , (1.10) is the weak formulation of the equation

$$(1.11) \quad \Delta g = \operatorname{div} f \quad \text{in } \Omega; \quad \int_{\Omega} g dx = 0; \quad \partial g / \partial \nu = 0 \quad \text{on } \partial \Omega.$$

To prove Proposition 1.1 we combine multiplier techniques, compactness arguments, Höllgren's Uniqueness Theorem and the following deep result that can be easily derived from the decoupling technique developed by D. Henry, O. Lopes and A. Perissinotto [HeLP].

**Theorem 1.2.** [HeLP]. Let  $P$  be as above and let us denote by  $\{S^0(t)\}_{t \geq 0}$  the strongly continuous semigroup in  $\mathcal{V}$  associated to the following decoupled system:

$$(1.12) \quad \begin{cases} u_{tt} - \mu \Delta u - (\lambda + \mu) \nabla \operatorname{div} u + \alpha \beta P u_t = 0 & \text{in } \Omega \times (0, \infty) \\ \theta_t - \Delta \theta + \beta \operatorname{div} u_t = 0 & \text{in } \Omega \times (0, \infty) \\ u = 0, \partial \theta / \partial \nu = 0 & \text{on } \Gamma \times (0, \infty) \\ u(0) = u^0, u_t(0) = u^1, \theta(0) = \theta^0 & \text{in } \Omega. \end{cases}$$

Then,  $S(t) - S^0(t) : \mathcal{V} \rightarrow C([0, T]; \mathcal{V})$  is continuous and compact (recall that  $S(t)$  is the semigroup associated to the system of thermoelasticity (1.1)).

We will denote by  $S_i^0(t)$ ,  $i = 1, 2, 3$  the three components of  $S^0(t)$ .

To our knowledge, the first results on controllability of thermoelastic systems are due to K. Narukawa [N]. But in [N] and in the more recent works by J. Lagnese [La], J. Lagnese and J. L. Lions [LaLi] and J. L. Lions [Li2], only "partial controllability" results are proved. Indeed, in these papers various models of thermoelasticity are considered but the results that are obtained are, roughly, that one may control exactly the displacement by means of one control acting in the equations of displacement when the coupling parameters are small enough. Note that nothing is said about the controllability of the temperature. More recently, S. Hansen [H] has proved the controllability of both the displacement and the temperature in one space dimension by means of one sole boundary-control, for various boundary conditions and without restrictions in the size of the coupling parameters. The methods of [H] are based on moment problems and nonharmonic Fourier series and they do not seem to extend to several space dimensions.

The results of [Z3] and of the present paper prove that in several space dimensions, without any restriction on the size of the coupling parameters, one may achieve simultaneously the exact controllability of the displacement and the approximate controllability of the temperature by means of one sole control.

The extension of these results to the case where both displacement and temperature satisfy Neumann boundary conditions is to be done.

The rest of the paper is organized as follows. In Section 2 we prove Theorem 1.1 following the method of [Z3]. In Section 4, we consider the case where the displacement satisfies Neumann boundary conditions and the temperature takes zero Dirichlet value.

## 2. Proof of Theorem 1.1

To prove Theorem 1.1 we need to use Proposition 1.1, Theorem 1.2 and a uniqueness or unique-continuation result for the adjoint system (1.8) as well. In the first three subsections we will state those results and give an outline of its prove. The proof of the Theorem 1.1 is concluded in the fourth subsection where the tools developed in [Li] and [FPZ1, 2, 3, 4] are applied.



### 2.1. A uniqueness result

**Proposition 2.1.** Suppose that  $T > \text{diam}(\Omega \setminus \omega)/\sqrt{\mu}$ . Let  $(\varphi, \psi)$  be a solution of system (1.8) such that  $\varphi = 0$  in  $\omega \times (0, T)$  and  $\int_{\Omega} \psi(x, t) dx = 0$  for all  $t \in [0, T]$ . Then  $\varphi \equiv \sigma \equiv 0$  in  $\Omega \times (0, T)$ .

This result is basically a consequence of Hölmgren's Uniqueness Theorem (see for instance [J]). Note however that we do not do any assumption about the set of zeros of the temperature  $\psi$ . But this is not necessary because of the very strong coupling between displacement and temperature. Indeed, since  $\varphi = 0$  in  $\omega \times (0, T)$ , then from the first equation in (1.8) we deduce that  $\nabla \psi_t = 0$  in  $\omega \times (0, T)$ . The arguments of [Z3] allow to conclude that  $\varphi \equiv 0$  and  $\psi \equiv c$  for some real constant  $c \in \mathbb{R}$ . The fact that  $c = 0$  is deduced from the hypothesis  $\int_{\Omega} \psi(x, t) dx = 0$ . For the technical details we refer to [Z3] where the case where both  $\varphi$  and  $\psi$  satisfy Dirichlet boundary conditions is treated in detail.

### 2.2. The observability inequality

This section is devoted to the proof of Proposition 1.1. We will use the notation  $\mathcal{V} = (L^2(\Omega))^n \times (H^{-1}(\Omega))^n \times \mathcal{H}$ .

First we observe that if  $(\varphi, \psi)$  solves (1.8), then  $(\phi, \psi)$ , where

$$\phi(x, t) = - \int_t^T \varphi(x, s) ds + \chi(x)$$

with

$$\begin{cases} -\mu \Delta \chi - (\lambda + \mu) \nabla \operatorname{div} \chi = -\varphi^1 - \beta \nabla \psi^0 & \text{in } \Omega \\ \chi = 0 & \text{on } \partial\Omega, \end{cases}$$

solves

$$(2.1) \quad \begin{cases} \phi_{tt} - \mu \Delta \phi - (\lambda + \mu) \nabla \operatorname{div} \phi + \beta \nabla \psi = 0 & \text{in } \Omega \times (0, T) \\ -\psi_t - \Delta \psi - \alpha \operatorname{div} \phi_t = 0 & \text{in } \Omega \times (0, T) \\ \phi = 0, \partial\psi/\partial\nu = 0 & \text{on } \partial\Omega \times (0, T) \\ \phi(x, T) = \chi(x), \phi_t(x, T) = \phi^0(x), \psi(x, T) = \psi^0(x) & \text{in } \Omega. \end{cases}$$

Taking into account that  $\|\chi\|_{(H_0^1(\Omega))^n}$  and  $\|\varphi^1 + \beta \nabla \psi^0\|_{(H^{-1}(\Omega))^n}$  are equivalent norms, we see that Proposition 1 is equivalent to the following one:

**Proposition 2.2.** Under the assumptions of Theorem 1.1, for every bounded set  $B$  of  $\mathcal{H}$  there exists  $\delta = \delta(B) > 0$  such that

$$(2.2) \quad \delta \leq \int_0^T \int_{\omega} |\phi_t|^2 dx dt$$

holds for every solution of (2.1) with initial data such that

$$(2.3) \quad \|(\chi, \varphi^0)\|_{(H_0^1(\Omega))^n \times (L^2(\Omega))^n} \geq 1, \psi^0 \in B.$$

Thus, it is sufficient to prove Proposition 2.2.

To prove Proposition 2.2, first we introduce the decoupled system associated to (2.1):

$$(2.4) \quad \begin{cases} \phi_t - \mu \Delta \phi - (\lambda + \mu) \nabla \operatorname{div} \phi - \alpha \beta P \phi_t = 0 & \text{in } \Omega \times (0, T) \\ \phi_t - \Delta \phi - \alpha \operatorname{div} \phi_t = 0 & \text{in } \Omega \times (0, T) \\ \phi = 0, \partial \phi / \partial \nu = 0 & \text{on } \partial \Omega \times (0, T) \\ \phi(x, T) = \chi(x), \phi_t(x, T) = \varphi^0(x), \psi(x, T) = \psi^0(x) & \text{in } \Omega \end{cases}$$

and consider the subsystem that  $\phi$  satisfies:

$$(2.5) \quad \begin{cases} \phi_t - \mu \Delta \phi - (\lambda + \mu) \nabla \operatorname{div} \phi - \alpha \beta P \phi_t = 0 & \text{in } \Omega \times (0, T) \\ \phi = 0 & \text{on } \partial \Omega \times (0, T) \\ \phi(x, T) = \chi(x), \phi_t(x, T) = \varphi^0(x) & \text{in } \Omega. \end{cases}$$

We have the following observability inequality for system (2.5).

**Proposition 2.3.** *Suppose that  $T > \operatorname{diam}(\Omega \setminus \omega) / \sqrt{\mu}$ . Then, there exists a constant  $C > 0$  and a semi-norm  $X : (H_0^1(\Omega))^n \times (L^2(\Omega))^n \rightarrow \mathbb{R}^+$  such that*

$$(2.6) \quad \|\chi\|_{(H_0^1(\Omega))^n}^2 + \|\varphi^0\|_{(L^2(\Omega))^n}^2 \leq C \left[ \int_0^T \int_{\omega} |\phi_t|^2 dx dt + X^2(\chi, \varphi^0) \right]$$

holds for every solution of (3.5),  $X : (H_0^1(\Omega))^n \times (L^2(\Omega))^n \rightarrow \mathbb{R}^+$  being continuous and compact.

**Proof.** The proof is the same as in Proposition 2.3 of [Z3]. Indeed, the only properties of the operator  $P$  that are used in [Z3] are the following ones:

$$\operatorname{curl} Pf = 0; \operatorname{div} Pf = \operatorname{div} f \quad \text{for all } f \in (L^2(\Omega))^n$$

and these conditions are fulfilled for  $P$  as defined in (1.10). ■

Let us now conclude the proof of Proposition 2.2.

We decompose the solution of (2.1) as  $(\phi, \psi) = (\tilde{\phi}, \tilde{\psi}) + (\xi, \eta)$  where  $(\tilde{\phi}, \tilde{\psi})$  solves (2.4) and  $(\xi, \eta)$  satisfies

$$(2.7) \quad \begin{cases} \xi_t - \mu \Delta \xi - (\lambda + \mu) \nabla \operatorname{div} \xi = -\alpha \beta P \tilde{\phi}_t - \beta \nabla \psi & \text{in } \Omega \times (0, T) \\ -\eta_t - \Delta \eta - \alpha \operatorname{div} \xi_t = 0 & \text{in } \Omega \times (0, T) \\ \xi = 0, \partial \eta / \partial \nu = 0 & \text{on } \partial \Omega \times (0, T) \\ \xi(x, T) = \xi_t(x, T) = 0, \eta(x, T) = 0 & \text{in } \Omega. \end{cases}$$

As a consequence of Proposition 2.3 we have

$$(2.8) \quad \|\chi\|_{(H_0^1(\Omega))^n}^2 + \|\varphi^0\|_{(L^2(\Omega))^n}^2 \leq C \left[ \int_0^T \int_{\omega} (|\phi_t|^2 + |\xi_t|^2) dx dt + X^2(\chi, \varphi^0) \right].$$

We argue by contradiction. Suppose that Proposition 2.2 does not hold. Then,

there exists a bounded set  $B$  of  $\mathcal{H}$  and a sequence of initial data  $(\chi_j, \psi_j) \in B$  satisfying (2.3) such that

$$(2.9) \quad \int_0^T \int_{\Omega} |\psi_j|^2 dx dt \rightarrow 0 \text{ as } j \rightarrow \infty.$$

In view of (2.8) and taking into account (2.9) and that

$$\|(\chi_j, \psi_j^0)\|_{H_0^1(\Omega) \times L^2(\Omega)} = 1$$

holds, we deduce that

$$(2.10) \quad \liminf_{j \rightarrow \infty} \int_0^T \int_{\Omega} |\xi_j|^2 dx dt + N^2(\chi_j, \psi_j^0) = 0.$$

We introduce the normalized data

$$(\chi_j^0, \psi_j^0, \psi_j^0) = (\chi_j^0, \psi_j^0, \psi_j^0) / \|(\xi_j, \psi_j^0, \psi_j^0)\|_{(L^2(\Omega))^3} + N^2(\chi_j, \psi_j^0)^{1/2}$$

and the corresponding solutions  $(\phi_j, \psi_j)$  and  $(\xi_j, \eta_j)$  of (2.1) and (2.7). We have then

$$(2.11) \quad \int_0^T \int_{\Omega} |\xi_j|^2 dx dt + N^2(\chi_j, \psi_j^0) = 1; \quad \int_0^T \int_{\Omega} |\phi_j|^2 dx dt \rightarrow 0 \text{ as } j \rightarrow \infty.$$

In view of (2.8) we deduce that

$$\|(\chi_j^0, \psi_j^0)\|_{H_0^1(\Omega) \times L^2(\Omega)} \leq C.$$

On the other hand, by (2.10) we know that  $(\psi_j^0)$  remains in a bounded set  $B$  of  $\mathcal{H}$ . By extracting subsequences we deduce that

$$\begin{cases} (\chi_j, \psi_j^0) \rightharpoonup (\chi, \psi^0) & \text{weakly in } (H_0^1(\Omega))^n \times (L^2(\Omega))^n \\ \psi_j^0 \rightharpoonup \psi^0 & \text{weakly in } L^2(\Omega) \end{cases}$$

and

$$(2.12) \quad \begin{cases} \psi_{j,t} \rightharpoonup \psi_t & \text{weakly in } (L^2(\Omega \times (0, T)))^n \\ \xi_{j,t} \rightharpoonup \xi_t & \text{weakly in } (L^2(\Omega \times (0, T)))^n \end{cases}$$

as  $j \rightarrow \infty$ , where  $(\hat{\varphi}, \hat{\psi})$  and  $(\xi, \eta)$  are the solutions of (1.8) and (2.7) corresponding to the limit initial data.

On the other hand, in virtue of Theorem 1.2 we know that  $(\hat{\xi}_{j,t})$  is relatively compact in  $C([0, T]; (L^2(\Omega))^n)$  and therefore

$$(2.13) \quad \hat{\xi}_{j,t} \rightarrow \hat{\xi}_t \text{ strongly in } (L^2(\Omega \times (0, T)))^n.$$

As a consequence of (2.11) and (2.12) we deduce that

$$(2.14) \quad \hat{\varphi} = \hat{\phi}_t = 0 \text{ in } \omega \times (0, T).$$

In view of (2.14) and applying Proposition 2.1 we obtain that  $(\varphi, \xi) = 0$  in  $\Omega \times (0, T)$  and therefore

$$(2.15) \quad (\varphi^0, \varphi^1, \varphi^0) = 0.$$

This implies that

$$(2.16) \quad \xi = 0.$$

However, combining (2.14), (2.13) and the fact that  $X : (H_0^1(\Omega))^n \times (L^2(\Omega))^n \rightarrow \mathfrak{H}^1$  is compact we deduce that

$$(2.17) \quad \|\xi\|_{L^2(\Omega \times (0, T))}^2 + X^2(\chi, \varphi^0) = 1$$

and this contradicts (2.15)-(2.16).

### 2.3. Proof of the decoupling result

In this subsection we give a sketch of the proof of Theorem 1.2 which follows the arguments of D. Henry, O. Lopes and A. Perissinotto [HeLP].

Let  $B$  be a bounded set of  $\mathcal{V}$ . We set

$$(u(t), u_t(t), \theta(t)) = [S(t)](u^0, u^1, \theta^0),$$

$$(\tilde{u}(t), \tilde{u}_t(t), \tilde{\theta}(t)) = [S^0(t)](u^0, u^1, \theta^0)$$

and

$$(v(t), v_t(t), \eta(t)) = [S(t) - S^0(t)](u^0, u^1, \theta^0)$$

for any  $(u^0, u^1, \theta^0) \in B$ .

We have:

$$(2.18) \quad \begin{cases} v_{tt} + \mu \Delta v + (\lambda + \mu) \nabla \operatorname{div} v + \alpha \nabla \eta = \alpha [\beta P \tilde{u}_t + \nabla \tilde{\theta}] & \text{in } \Omega \times (0, T) \\ \eta_{tt} + \Delta \eta + \beta \operatorname{div} v_t = 0 & \text{in } \Omega \times (0, T) \\ v = 0, \partial \eta / \partial \nu = 0 & \text{on } \Gamma \times (0, T) \\ v(0) = v_t(0) = 0, \eta(0) = 0 & \text{in } \Omega. \end{cases}$$

It is sufficient to check that  $\beta P \tilde{u}_t + \nabla \tilde{\theta}$  is bounded in  $L^1(0, T; (H^s(\Omega))^n)$  for some  $s > 0$  when  $(u^0, u^1, \theta^0)$  varies in  $B$ .

Let us decompose  $\beta P \tilde{u}_t + \nabla \tilde{\theta}$  as follows:

$$\beta P \tilde{u}_t + \nabla \tilde{\theta} = \nabla w_1 + \nabla w_2$$

where  $w_1$  satisfies

$$(2.19) \quad \begin{cases} w_{1,t} + \Delta w_1 = 0 & \text{in } \Omega \times (0, T) \\ \partial w_1 / \partial \nu = 0 & \text{on } \Gamma \times (0, T) \\ w_1(0) = \beta g + \theta^0 & \text{in } \Omega \end{cases}$$

with  $g \in V$  the solution of (1.10) for  $f = u^1$  and  $w_2$  verifies

$$(2.20) \quad \begin{cases} w_{2,t} - \Delta w_2 = \beta h_H & \text{in } \Omega \times (0, T) \\ \partial w_2 / \partial \nu = 0 & \text{on } \Gamma \times (0, T) \\ w_2(0) = 0 & \text{in } \Omega \end{cases}$$

where, for each  $t \in [0, T]$ ,  $h(x, t)$  solves (1.10) with  $f(x) = \dot{u}(x, t)$ .

Since  $\beta g = \theta^0$  is bounded in  $L^2(\Omega)$ , because of the regularizing effect of the heat equation (2.19), we deduce that  $w_1$  is bounded in  $L^1(0, T; H^{1+s}(\Omega))$  for any  $0 < s < 1$ .

On the other hand, by elliptic regularity it is easy to check that  $h_H$  is bounded in  $L^\infty(0, T; L^2(\Omega))$ . Therefore,  $w_2$  is bounded in  $L^1(0, T; H^2(\Omega))$ . This concludes the proof of Theorem 1.2.

## 2.4. Conclusion

In this subsection we conclude the proof of Theorem 1.1.

First we observe that it is sufficient to consider the case where  $u^0 = u^1 = 0$  and  $\theta^0 \equiv 0$ . Therefore, in the sequel we will assume that  $u^0 \equiv u^1 \equiv 0$  and  $\theta^0 \equiv 0$ .

Given any  $(v^0, v^1, \eta^0) \in \mathcal{V}$  and  $\varepsilon > 0$  we introduce the functional  $J : \mathcal{V} \rightarrow \mathbb{R}$  defined as follows

$$(2.21) \quad J(\varphi^0, \varphi^1, \psi^0) = \frac{1}{2} \int_0^T \int_\Omega |\varphi|^2 dx dt + \int_\Omega v^1 \cdot \varphi^0 dx + \langle v^0, \varphi^1 \rangle + \|\psi^0\|_{L^2(\Omega)} + \int_\Omega (\eta^0 + \beta \operatorname{div} v^0) \psi^0 dx$$

where  $\langle \cdot, \cdot \rangle$  denotes the duality product between  $(H_0^1(\Omega))^n$  and  $(H^{-1}(\Omega))^n$  and  $(\varphi, \psi)$  the solution of (1.8).

The functional  $J$  is coercive in  $\tilde{\mathcal{V}}$ . More precisely, we have the following result:

**Lemma 2.1.** *Under the assumptions of Theorem 1.1,*

$$(2.22) \quad \liminf_{\|(v^0, v^1) + \beta \nabla \psi^0\|_{\mathcal{V}} \rightarrow +\infty} J(\varphi^0, \varphi^1, \psi^0) = +\infty.$$

In view of Proposition 1.1, the proof of this lemma can be done following the arguments of Lemma 4.1 in [Z1] in a straightforward way.

By using the coercivity property (2.22) it is easy to check that the infimum of  $J$  over  $\tilde{\mathcal{V}}$  is achieved at some  $(\hat{\varphi}^0, \hat{\varphi}^1, \hat{\psi}^0) \in \tilde{\mathcal{V}}$  (see [FPZ1, 2, 3, 4] and [Li] for the details and some other related questions). At this minimizer  $(\hat{\varphi}^0, \hat{\varphi}^1, \hat{\psi}^0)$  we have the following optimality conditions:

$$(2.23) \quad \left| \int_0^T \int_\Omega \hat{\varphi} \cdot \rho dx dt - \int_\Omega v^1 \cdot \rho^0 dx + \langle v^0, \rho^1 \rangle - \int_\Omega (\eta^0 + \beta \operatorname{div} v^0) \xi^0 dx \right| \leq \varepsilon \|\xi^0\|_{L^2(\Omega)}$$

for all  $(\rho^0, \rho^1, \xi^0) \in \mathcal{V}$  where  $(\bar{\varphi}, c)$  denotes the solution of (1.8) corresponding to the minimizer  $(\bar{\varphi}^0, \bar{\varphi}^1, c^0)$  and  $(\rho, \xi)$  the solution of (1.8) with data  $(\rho^0, \rho^1, \xi^0)$ .

Observe that if  $(u, \theta)$  solves (1.1) with  $u^0 = u^1 = 0$  and  $\theta^0 = 0$  and  $f = \bar{\varphi}$  then

$$(2.24) \quad \int_0^T \int_{\Omega} \bar{\varphi} \cdot \rho dx dt = \int_{\Omega} u_t(T) \cdot \rho^0 dx = \langle u(T), \rho^1 \rangle + \int_{\Omega} (\theta(T) + t \operatorname{div} u(T)) \xi^0 dx.$$

Combining (2.23) and (2.24) we easily deduce that  $(u, \theta)$  satisfy (1.3).

### 3. Boundary conditions of the second type

When the displacement satisfies Neumann boundary conditions and the temperature takes Dirichlet boundary conditions the controlled system of thermoelasticity reads as follows:

$$(3.1) \quad \begin{cases} u_{tt} - \mu \Delta u + (\lambda + \mu) \nabla \operatorname{div} u + \alpha \nabla \theta = f \chi_{\omega} & \text{in } \Omega \times (0, \infty) \\ \theta_t - \Delta \theta + t \operatorname{div} u_t = 0 & \text{in } \Omega \times (0, \infty) \\ \mu \partial u / \partial \nu + (\lambda + \mu) \operatorname{div} u \cdot \nu = 0, \quad \theta = 0 & \text{on } \Gamma \times (0, \infty) \\ u(x, 0) = u^0(x), u_t(x, 0) = u^1(x), \theta(x, 0) = \theta^0(x) & \text{in } \Omega. \end{cases}$$

In this situation we have the following controllability result:

**Theorem 3.1.** *Let  $\omega$  be a neighborhood of the boundary  $\Gamma$  in  $\Omega$ . Suppose that  $T > \operatorname{diam}(\Omega \setminus \omega) / \sqrt{\mu}$ . Then, system (3.1) is exact-approximately controllable in time  $T$  in the following sense: for any  $(u^0, u^1, \theta^0), (v^0, v^1, \eta^0) \in \mathcal{V}_1 = (H^1(\Omega))^n \times (L^2(\Omega))^n \times L^2(\Omega)$  and  $\varepsilon > 0$  there exists a control function  $f \in (L^2(\Omega \times (0, T)))^n$  such that the solution of (3.1) satisfies (1.3).*

Note that, in this case, there is no restriction like (1.5) on the final data we may reach starting from a given initial data. This is due to the fact that there is no quantity that remains constant in time, this constant being independent of  $f$ .

The proof of this theorem is similar to that of Theorem 1.1. However, in this case, the decoupling operator  $P$  has to be defined in a different way. In this case,  $P : (L^2(\Omega))^n \times (L^2(\Omega))^n$  is given by  $Pf = \nabla g$  where  $g \in H_0^1(\Omega)$  is the solution of

$$(3.2) \quad \int_{\Omega} \nabla g \cdot \nabla \rho dx = \int_{\Omega} f \cdot \nabla \rho dx, \quad \text{for all } \rho \in H_0^1(\Omega).$$

Obviously, the solution  $g$  of (3.2) in  $H^1(\Omega) \cap \mathcal{H}$  exists and it is unique. Note that (3.2) is the weak formulation of the equation

$$(3.3) \quad \Delta g = \operatorname{div} f \quad \text{in } \Omega; g = 0 \quad \text{on } \partial\Omega.$$

### Acknowledgements.

This work was done while the author was visiting the Department of Computational and Applied Mathematics of Rice University in Houston with the support of the "Ministerio de Educación y Ciencia (Spain)". This research was partially supported by DGICYT Project PB90-0245 and by the EEC Grant SC1\*-CT91-0732.

### References

- [FPZ1] C. Fabre, J. P. Puel and E. Zuazua, Contrôlabilité approchée de l'équation de la chaleur semilinéaire, *C. R. Acad. Sci. Paris*, 315, série I, 1992, 807-812.
- [FPZ2] C. Fabre, J. P. Puel and E. Zuazua, Approximate controllability of the semilinear heat equation, *Proc. Royal Soc. Edinburgh*, to appear.
- [FPZ3] C. Fabre, J. P. Puel and E. Zuazua, Contrôlabilité approchée de l'équation de la chaleur linéaire avec des contrôles de norme  $L^\infty$  minimale, *C. R. Acad. Sci. Paris*, 316, série I, 1993, 679-684.
- [FPZE] C. Fabre, J. P. Puel and E. Zuazua, On the density of the range of the semigroup for semilinear heat equations, *IMA Preprint Series*, # 1093, 1992.
- [H] S. W. Hansen, Boundary control of a one-dimensional, linear, thermoelastic rod, *SIAM J. Cont. Optim.*, to appear.
- [HeLP] D. Henry, O. Lopes and A. Perissinotto, On the essential spectrum of a semigroup of thermoelasticity, preprint.
- [J] F. John, *Partial Differential Equations*, Applied Mathematical Sciences 1, 4th edition, Springer-Verlag, New York, 1986.
- [La] J. E. Lagnese, The reachability problem for thermoelastic plates, *Arch. Rat. Mech. Anal.*, 112 (1990), 223-267.
- [LaLi] J. E. Lagnese and J. L. Lions, *Modelling, Analysis and Control of Thin Plates*, Masson, RMA 6, Paris, 1988.
- [Li] J. L. Lions, Remarks on approximate controllability, *J. Analyse. Math.*, 59, 1992, 103-116.
- [N] K. Nakakawa, Boundary value control of thermoelastic systems, *Hiroshima Math. J.*, 13 (1983), 227-272.
- [Z1] E. Zuazua, Contrôlabilité du système de la thermoélasticité, *C. R. Acad. Sci. Paris*, 317 (1993), 371-376.
- [Z2] E. Zuazua, Contrôlabilité du système de la thermoélasticité sans restrictions sur les paramètres de couplage, *C. R. Acad. Sci. Paris*, 318 (1994), 643-648.
- [Z3] E. Zuazua, Controllability of the linear system of thermoelasticity, preprint.

Enrique Zuazua  
 Departamento de Matemática Aplicada  
 Universidad Complutense  
 28040 Madrid, Spain

In order to be consistent with our basic kinematic hypothesis, it is assumed that  $\mathbf{F}$  and  $\rho$  are independent of  $x_3$ . Let us further assume that the material in question is *linearly elastic* (hookian), i.e.,

$$(1.7) \quad s^{ij} = C^{ijkl} \varepsilon_{kl},$$

where the coefficients of elasticity  $C^{ijkl}$  depend only on  $x_1$  and  $x_2$ . The symmetry of  $s^{ij}$  and  $\varepsilon_{kl}$  requires that

$$(1.8) \quad C^{ijkl} = C^{jikl} = C^{ijlk},$$

and we further assume that

$$(1.9) \quad C^{ijkl} = C^{klij}$$

Then, in particular, the stresses  $s^{ij}$  do not depend on  $x_3$  and (1.6) reduces to

$$(1.10) \quad s_{,\alpha}^{ij} \mathbf{a}_j + (s^{\alpha\beta} \mathbf{W}_{,\beta})_{,\alpha} + \mathbf{F} = \rho \mathbf{W}_{,\alpha}.$$

This system comprises three scalar equations for the scalar components  $W_i$  of  $\mathbf{W}$ .

**1.2. Edge Conditions.** Let  $\Lambda$  denote the two-dimensional boundary of the reference cross-section  $\mathcal{P}$  and assume that

$$(1.11) \quad \Lambda = \bar{\Lambda}^C \cup \bar{\Lambda}^D,$$

where  $\Lambda^C, \Lambda^D$  are mutually disjoint, relatively open subsets of  $\Lambda$  and  $\Lambda^D \neq \emptyset$ . (The superscripts  $C$  and  $D$  are used to indicate "clamped" and "dynamic," respectively.) The partition (1.11) induces a similar partition of  $\Gamma := \partial\Omega$  into

$$(1.12) \quad \Gamma = \bar{\Gamma}^C \cup \bar{\Gamma}^D$$

through the mapping

$$\Gamma \mapsto \Lambda : (x_1, x_2) \mapsto \mathbf{p}_0 + x_\alpha \mathbf{a}_\alpha.$$

Let  $(\nu_1, \nu_2)$  denote a unit exterior normal vector and  $(-\nu_2, \nu_1)$  a unit tangent vector at a point  $(x_1, x_2)$  of  $\Gamma$ . Since the mapping  $(x_1, x_2, x_3) \mapsto x_i \mathbf{a}_i$  is orthogonal with determinant unity,  $\boldsymbol{\nu} := \nu_1 \mathbf{a}_1 + \nu_2 \mathbf{a}_2$  and  $\boldsymbol{\tau} := -\nu_2 \mathbf{a}_1 + \nu_1 \mathbf{a}_2$  are vectors which are normal and tangent, respectively, to  $\Lambda$  at the point  $\mathbf{p}_0 + x_\alpha \mathbf{a}_\alpha$  with  $\boldsymbol{\nu}$  pointing out of  $\mathcal{P}$  and  $\boldsymbol{\tau}$  positively oriented with respect to  $\mathbf{a}_3$ , i.e.,  $\boldsymbol{\tau} \times \mathbf{a}_3 = \boldsymbol{\nu}$ .

The edge conditions along the part of  $\partial\mathcal{B}$  corresponding to  $\Lambda^C$  are described by

$$(1.13) \quad \mathbf{W} = 0 \quad \text{on } \Gamma^C, \quad t > 0.$$

Let  $\mathbf{f}$  denote a given distribution of forces along  $\Lambda^D + x_3 \mathbf{a}_3$ ,  $-\infty < x_3 < +\infty$  which is independent of  $x_3$ . The condition for the balance of forces along this part of  $\partial\mathcal{B}$  is

$$\nu_\alpha \mathbf{t}^\alpha = \mathbf{f}$$



which, in view of our previous hypotheses, may be written

$$(1.14) \quad \nu_\alpha s^{\alpha i} \mathbf{a}_i + \nu_\alpha s^{\alpha i} \mathbf{W}_{i\alpha} = \mathbf{f} \text{ on } \Gamma^D, t \geq 0.$$

**1.3. Hamilton's Principle.** Let  $\mathbf{W}$  satisfy (1.10) and (1.14), and let  $\mathbf{W} = W_i \mathbf{a}_i$  be a smooth function defined on  $\Omega \times [0, T]$  such that  $\mathbf{W}_{i\alpha}|_{t=0} = \mathbf{W}_{i\alpha}|_{t=T} = 0$ ,  $\mathbf{W}|_{\Gamma^D \times (0, T)} = 0$ . We form the scalar product in  $\mathcal{H}^1$  of (1.10) with  $\mathbf{W}$  and integrate over  $\Omega \times (0, T)$ . After some integrations by parts we obtain

$$(1.15) \quad \int_0^T \int_\Omega (\rho \mathbf{W}_t \cdot \mathbf{W}_t - s^{\alpha\beta} W_{i\alpha} \cdot s^{\alpha\beta} \mathbf{W}_{i\beta} \cdot \mathbf{W}_{i\alpha}) d\Omega \\ + \int_0^T \int_\Omega \mathbf{F} \cdot \mathbf{W} d\Omega + \int_0^T \int_\Gamma \mathbf{f} \cdot \mathbf{W} d\Gamma = 0.$$

Conversely, if  $\mathbf{W}$  satisfies (1.15) for all such test functions  $\mathbf{W}$ , then  $\mathbf{W}$  satisfies (in an appropriate sense) (1.10) and (1.14). The variational equation (1.15) is known as the *principle of virtual work* for the membrane problem under consideration.

We now recast (1.15) into a form known as *Hamilton's principle*. Let  $\delta$  denote the first Frechet derivative with respect to  $\mathbf{W}$ . We may write

$$(1.16) \quad s^{\alpha\beta} W_{i\alpha} \cdot s^{\alpha\beta} \mathbf{W}_{i\beta} \cdot \mathbf{W}_{i\alpha} = s^{\alpha\beta} \delta_{i\alpha\beta} \mathbf{W}_{i\beta} \cdot \mathbf{W}_{i\alpha}$$

where  $\delta_{i\alpha\beta} \mathbf{W}_{i\beta} \cdot \mathbf{W}_{i\alpha}$  denotes the value of  $\delta_{i\alpha\beta}$  at  $\mathbf{W}$ . For any material obeying (1.7) (1.9) one has

$$(1.17) \quad s^{\alpha\beta} \delta_{i\alpha\beta} = C^{ijkl}{}_{\alpha\beta} \delta_{i\alpha} \delta_{j\beta} = \frac{1}{2} \delta(C^{ijkl}{}_{\alpha\beta}) = \frac{1}{2} \delta(s^{\alpha\beta}{}_{ij}).$$

Under assumption (1.1),

$$(1.18) \quad s^{\alpha\beta} \delta_{i\alpha\beta} = s^{\alpha\beta}{}_{i\alpha\beta} + 2s^{\alpha\beta}{}_{i\alpha\alpha}.$$

It then follows from (1.16)–(1.18) that (1.15) may be written

$$(1.19) \quad \delta \int_0^T \lambda(t) - \mathcal{U}(t) + \mathcal{W}(t) dt = 0$$

where

$$(1.20) \quad \lambda(t) = \frac{1}{2} \int_\Omega \rho |\mathbf{W}_t|^2 d\Omega$$

is the *kinetic energy* of the deformation,

$$(1.21) \quad \mathcal{U}(t) = \frac{1}{2} \int_\Omega (s^{\alpha\beta}{}_{i\alpha\beta} + 2s^{\alpha\beta}{}_{i\alpha\alpha}) d\Omega \\ + \frac{1}{2} \int_\Omega (s^{\alpha\beta} W_{i\alpha} \cdot \frac{1}{2} s^{\alpha\beta} \mathbf{W}_{i\beta} \cdot \mathbf{W}_{i\alpha}) d\Omega$$

is the *strain energy*, and

$$(1.22) \quad W(t) = \int_{\Omega} \mathbf{F} \cdot \mathbf{W} \, d\Omega + \int_{\Gamma} \mathbf{f} \cdot \mathbf{W} \, d\Gamma$$

is the *work of external forces*. The variational formulation (1.19) with  $\mathbf{A}$ ,  $\mathcal{H}$  and  $W$  so defined, is Hamilton's principle for the membrane problem under consideration.

## 2. Systems of Interconnected Elastic Membranes

For  $i = 1, \dots, n$ , let  $\mathbf{a}_1^i, \mathbf{a}_2^i, \mathbf{a}_3^i$  be a right-handed orthonormal system in  $\mathbb{R}^3$ , let  $\mathbf{p}_0^i$  be a fixed vector in  $\mathbb{R}^3$ , and  $\Omega_i$  be a bounded, open, connected region in  $\mathbb{R}^2$  with Lipschitz boundary consisting of a finite number of smooth simple curves. We set

$$(2.1) \quad \begin{aligned} \mathbf{p}_i(x_1, x_2) &= \mathbf{p}_0^i + \sum_{\alpha} x_{\alpha} \mathbf{a}_{\alpha}^i, \\ \mathbf{r}_i(x_1, x_2, x_3) &= \mathbf{p}_i(x_1, x_2) + x_3 \mathbf{a}_3^i, \\ \mathcal{P}_i &= \{\mathbf{p}_i(x_1, x_2) \mid (x_1, x_2) \in \Omega_i\}. \end{aligned}$$

The mapping  $(x_1, x_2) \mapsto \mathbf{p}_i(x_1, x_2)$  is a homeomorphism of  $\bar{\Omega}_i$  onto  $\bar{\mathcal{P}}_i$  and of  $\partial\Omega_i$  onto  $\partial\mathcal{P}_i$ , with

$$\mathbf{p}_i^{-1}(\mathbf{p}) = ((\mathbf{p} - \mathbf{p}_0^i) \cdot \mathbf{a}_1^i, (\mathbf{p} - \mathbf{p}_0^i) \cdot \mathbf{a}_2^i).$$

**Remark 2.1.** As in the last section, Greek indices take values 1, 2, but, unlike before, we *do not* assume summation convention and Roman indices are not necessarily restricted to set  $\{1, 2, 3\}$ . The range of such an index will be explicitly indicated whenever it is not clear from context. We do, however, make the convention that

$$\sum_i = \sum_{i=1}^3.$$

The set  $\mathcal{P}_i$  is considered as a reference cross-section of an infinite cylinder whose reference configuration is

$$\mathcal{B}_i = \{\mathbf{r}_i(x_1, x_2, x_3) \mid (x_1, x_2) \in \Omega_i, |x_3| < \infty\}.$$

We make the following hypotheses regarding the membranes  $\mathcal{P}_i$ .

$$\mathcal{P}_i \cap \mathcal{P}_k = \emptyset, \quad \forall i \neq k;$$

$$U_i^0 \cap \bar{\mathcal{P}}_i \text{ is a connected set in } \mathbb{R}^3;$$

$$\mathcal{P}_i \cap \mathcal{P}_k \text{ is either empty or is a finite union of components, } \forall i \neq k.$$

Each of the bodies  $\mathcal{B}_i$  is assumed to satisfy the basic kinematic assumption (1.1): the position vector  $\mathbf{R}'(x_1, x_2, x_3)$  to the displaced particle originally at  $\mathbf{r}_i(x_1, x_2, x_3)$  is given by

$$(2.2) \quad \mathbf{R}'(x_1, x_2, x_3, t) = \mathbf{r}'(x_1, x_2, x_3) + \mathbf{W}'(x_1, x_2, t), \\ (x_1, x_2) \in \Omega_i, |x_3| < \infty, i = 1, \dots, n.$$

**2.1. Geometric Junction Conditions.** We tentatively define a *joint*, or *junction*, of the multi-membrane system as a component (a maximal line segment; by definition, line segments are connected) which is a common edge of two or more reference cross-sections, that is, a maximal line segment where a  $\bar{\mathcal{P}}_i$  and  $\bar{\mathcal{P}}_k$  intersect for some  $k \neq i$ . However, this definition allows the possibility that the set of all such components does not form a mutually disjoint collection. If  $C_1$  and  $C_2$  are components such that  $C_1 \cap C_2 \neq \emptyset$ , we replace  $C_1$  and  $C_2$  by  $C_1 \cup C_2$ ,  $C_1 / (C_1 \cap C_2)$  and  $C_2 / (C_1 \cap C_2)$ , whose union is  $C_1 \cup C_2$ . Each set  $C'_i = (C_1 \cup C_2)$  is either a single line segment or the union of disjoint segments. In the latter case we consider  $C'_i / (C_1 \cap C_2)$  to consist of several separate components. In this manner one may eliminate the overlapping of components. Finally, we define the joints of the system to be the collection of disjoint, open line segments which are the *interners* of the collection of non-overlapping components.

For any joint  $J$  of the multi-membrane system define

$$\mathcal{I}(J) = \{i \in [1, \dots, n] \mid J \subset \partial\mathcal{P}_i\}.$$

Thus  $\mathcal{I}(J)$  is the index set of those  $\mathcal{P}_i$  which share  $J$ . The deformation of the overall system is constrained by the following *geometric junction condition* at each joint  $J$ :

$$(2.3) \quad \mathbf{W}'(\mathbf{p}_i^{-1}(\mathbf{p}), t) = \mathbf{W}'(\mathbf{p}_j^{-1}(\mathbf{p}), t), \quad \mathbf{p} \in J, i, j \in \mathcal{I}(J).$$

Condition (2.3) simply requires that each membrane sharing  $J$  have the same displacement there.

**2.2. Dynamic Conditions.** The geometric joint conditions reflect the continuity of the overall structure. In addition, there are *dynamic junction conditions* which represent the balance of linear momentum at the joints. The simplest way to obtain these is to minimize the Hamiltonian associated with the structure. In carrying out the calculation, the variation must be taken with respect to displacements which satisfy the geometric boundary conditions and the geometric joint conditions. Of course, in addition to the dynamic junction conditions Hamilton's principle also yields the equations of motion and dynamic boundary conditions of the structure, as usual.

Set  $\Lambda_i = \partial\mathcal{P}_i$ . We assume that

$$(2.4) \quad \Lambda_i = \bar{\Lambda}_i^C \cup \bar{\Lambda}_i^D \cup \bar{\Lambda}_i^J,$$

where  $\Lambda_i^C$ ,  $\Lambda_i^D$ ,  $\Lambda_i^J$  are disjoint and relatively open in  $\Lambda_i$ . This partition of  $\Lambda_i$  induces a similar partition of  $\Gamma_i := \partial\Omega_i$  into

$$(2.5) \quad \Gamma_i = \bar{\Gamma}_i^C \cup \bar{\Gamma}_i^D \cup \bar{\Gamma}_i^J,$$

where

$$\Lambda_i^C = \mathbf{p}_i(\Gamma_i^C), \quad \Lambda_i^D = \mathbf{p}_i(\Gamma_i^D), \quad \Lambda_i^J = \mathbf{p}_i(\Gamma_i^J).$$

The set  $\Lambda_i^C$  represents the part of  $\Lambda_i$  that is clamped, i.e.,

$$(2.6) \quad \mathbf{W}^i = 0 \quad \text{on } \Gamma_i^C,$$

$\Lambda_i^D$  is the part of  $\Lambda_i$  along which "dynamic" boundary conditions will hold, and  $\Lambda_i^J$  is that part of  $\Lambda_i$  such that

$$\Lambda_i^J \cap \Lambda_k \neq \emptyset \quad \text{for at least one } k \neq i.$$

Thus  $\Lambda_i^J$  consists of those joints of the multi-membrane system which belong to  $\Lambda_i$ . We label the entire collection of joints by  $J_1, J_2, \dots, J_m$ .

The total kinetic energy of the system is defined to be

$$\mathcal{K}(t) = \frac{1}{2} \sum_{i=1}^n \int_{\Omega_i} \rho_i |\mathbf{W}_{i,t}^i|^2 d\Omega,$$

where  $\rho_i(x_1, x_2)$  is the mass density per unit of reference area of  $\mathcal{P}_i$ . Let  $s_i^{jk}$  denote the stress tensor associated with the  $i$ th membrane. The total strain energy is defined to be

$$\mathcal{U}(t) = \frac{1}{2} \sum_{i=1}^n \int_{\Omega_i} \left( \sum_{\alpha, \beta} s_i^{\alpha\beta} W_{i,\alpha}^i W_{i,\beta}^i + \frac{1}{2} \sum_{\alpha, \beta} s_i^{\alpha\beta} \mathbf{W}_{i,\alpha}^i \cdot \mathbf{W}_{i,\beta}^i \right) d\Omega.$$

Let  $\mathbf{F}^i$  denote an external body force acting on the  $i$ th membrane  $\mathcal{P}_i$ ,  $\mathbf{f}^i$  be a force acting along  $\Lambda_i^D$  and  $\tilde{\mathbf{f}}^k$  be an applied force along the junction region  $J_k$ . The work of these external forces is defined to be

$$\begin{aligned} \mathcal{W}(t) &= \sum_{i=1}^n \left\{ \int_{\mathcal{P}_i} \mathbf{F}^i \cdot \mathbf{W}^i d\Lambda_i + \int_{\Lambda_i^D} \mathbf{f}^i \cdot \mathbf{W}^i d\Lambda \right\} + \sum_{k=1}^m \int_{J_k} \tilde{\mathbf{f}}^k \cdot \mathbf{W}^{i_k} dJ_k \\ &= \sum_{i=1}^n \left\{ \int_{\Omega_i} \mathbf{F}^i \cdot \mathbf{W}^i d\Omega_i + \int_{\Gamma_i^D} \mathbf{f}^i \cdot \mathbf{W}^i d\Gamma \right\} + \sum_{k=1}^m \int_{\Gamma_{i_k}(J_k)} \tilde{\mathbf{f}}^k \cdot \mathbf{W}^{i_k} d\Gamma, \end{aligned}$$

where

$$\begin{aligned} \Gamma_i(J_k) &= \{(x_1, x_2) \in \Gamma_i^J \mid \mathbf{p}_i(x_1, x_2) \in J_k\} = \{\mathbf{p}_i^{-1}(\mathbf{p}) \mid \mathbf{p} \in J_k\}, \\ i &= 1, \dots, n; \quad k = 1, \dots, m, \end{aligned}$$

and where we may take

$$i_k = \min\{i \mid i \in \mathcal{I}(J_k)\}$$

in accordance with the geometric junction conditions (2.3). In the integrals over  $\mathcal{P}_i$  and  $\Lambda_i^D$ , the integrands are evaluated at  $\mathbf{p}_i^{-1}(\mathbf{p})$  with  $\mathbf{p} \in \mathcal{P}_i$  and  $\mathbf{p} \in \Lambda_i^D$ , respectively, while the integrand in the integral over  $J_k$  is evaluated at  $\mathbf{p}_{ik}^{-1}(\mathbf{p})$ ,  $\mathbf{p} \in J_k$ . The dynamic equations of the multi-membrane system are *postulated* to be those associated with the Hamilton's principle

$$(2.7) \quad \delta \int_0^T [\mathcal{K}(t) - \mathcal{U}(t) + \mathcal{W}(t)] dt = 0,$$

in which the variation  $\delta$  is taken with respect to displacements  $(\mathbf{W}^1, \dots, \mathbf{W}^n)$  which respect the geometric boundary conditions (2.6) and geometric junction conditions (2.3) along each joint  $J_1, \dots, J_m$ . It is fairly obvious that we will obtain, in particular, the equations of motion (1.10) for each membrane of the system and the dynamic boundary conditions (1.14) along  $\Gamma_i^D$ . Thus we have for each  $i = 1, \dots, n$  the equations of motion

$$(2.8) \quad \sum_{\alpha, j} s_{i, \alpha}^{\alpha, j} \mathbf{a}'_j + \sum_{\alpha, j} (s_{i, \alpha}^{\alpha, j} \mathbf{W}'_{j, \alpha})_{, \alpha} + \mathbf{F}' = \rho_i \mathbf{W}'_{, tt}, \quad (x_1, x_2) \in \Omega_i, \quad t > 0,$$

and the boundary conditions

$$(2.9) \quad \sum_{\alpha, j} \nu'_\alpha s_{i, \alpha}^{\alpha, j} \mathbf{a}'_j + \sum_{\alpha, j} \nu'_\alpha s_{i, \alpha}^{\alpha, j} \mathbf{W}'_{j, \alpha} = \mathbf{f}' \quad \text{on } \Gamma_i^D, \quad t > 0.$$

There are, in addition, junction conditions given by the following variational equation:

$$(2.10) \quad \sum_{i=1}^n \int_{\Gamma_i^I} \left( \sum_{\alpha, j} \nu'_\alpha s_{i, \alpha}^{\alpha, j} \mathbf{a}'_j + \sum_{\alpha, j} \nu'_\alpha s_{i, \alpha}^{\alpha, j} \mathbf{W}'_{j, \alpha} \right) \cdot \dot{\mathbf{W}}^i d\Gamma = \sum_{k=1}^m \int_{J_k} \mathbf{f}^k \cdot \mathbf{W}^k dJ_k$$

for all sufficiently smooth displacements  $\mathbf{W}^i$ ,  $i = 1, \dots, n$ , which satisfy the geometric boundary and junction conditions.

We have

$$\sum_{i=1}^n \int_{\Gamma_i^I} = \sum_{k=1}^m \sum_{i \in I(J_k)} \int_{\Gamma_i(J_k)}.$$

The segment  $J_k$  may be parameterized as

$$J_k = \{\mathbf{q}_k^0 + s\boldsymbol{\tau}^k \mid a_k \leq s \leq b_k\},$$

where  $\mathbf{q}_k^0 \in J_k$  is fixed but arbitrary. Since  $J_k \subset \mathbf{p}_i(\Gamma_i(J_k))$ , this parameterization of  $J_k$  induces a parameterization of the segment  $\Gamma_i(J_k)$  given by

$$\Gamma_i(J_k) = \{\mathbf{x}_{ik}^0 + s(\tau_1, \tau_2) \mid a_k \leq s \leq b_k\}.$$

where  $\tau^{ik} = \tau_1 \mathbf{a}_1^{ik} + \tau_2 \mathbf{a}_2^{ik}$  and  $\mathbf{x}_{ik}^0$  is some point of  $\Gamma_i(J_k)$  depending on both  $i$  and  $k$ . Therefore

$$\begin{aligned} \int_{\Gamma_i(J_k)} \left( \sum_{\alpha,j} \nu_\alpha^i s_i^{\alpha,j} \mathbf{a}_j^i + \sum_{\alpha,\beta} \nu_\alpha^i s_i^{\alpha,\beta} \mathbf{W}_{,\beta}^i \right) \cdot \hat{\mathbf{W}}^i(\mathbf{x}) d\Gamma \\ = \int_{a_k}^{b_k} \left( \sum_{\alpha,j} \nu_\alpha^i s_i^{\alpha,j} \mathbf{a}_j^i + \sum_{\alpha,\beta} \nu_\alpha^i s_i^{\alpha,\beta} \mathbf{W}_{,\beta}^i \right) \cdot \hat{\mathbf{W}}^i(\mathbf{x}) ds \\ \quad (\mathbf{x} = \mathbf{x}_{ik}^0 + s(\tau_1, \tau_2)) \\ = \int_{a_k}^{b_k} \left( \sum_{\alpha,j} \nu_\alpha^i s_i^{\alpha,j} \mathbf{a}_j^i + \sum_{\alpha,\beta} \nu_\alpha^i s_i^{\alpha,\beta} \mathbf{W}_{,\beta}^i \right) \cdot \mathbf{W}^i(\mathbf{p}_i^{-1}(\mathbf{p})) ds \\ \quad (\mathbf{p} = \mathbf{q}_k^0 + s\tau^{ik}) \\ = \int_{J_k} \left( \sum_{\alpha,j} \nu_\alpha^i s_i^{\alpha,j} \mathbf{a}_j^i + \sum_{\alpha,\beta} \nu_\alpha^i s_i^{\alpha,\beta} \mathbf{W}_{,\beta}^i \right) \cdot \hat{\mathbf{W}}^i(\mathbf{p}_i^{-1}(\mathbf{p})) dJ_k. \end{aligned}$$

Thus (2.10) may be expressed as

$$\begin{aligned} \sum_{k=1}^m \sum_{i \in I(J_k)} \int_{J_k} \left( \sum_{\alpha,j} \nu_\alpha^i s_i^{\alpha,j} \mathbf{a}_j^i + \sum_{\alpha,\beta} \nu_\alpha^i s_i^{\alpha,\beta} \mathbf{W}_{,\beta}^i \right) \cdot \hat{\mathbf{W}}^i(\mathbf{p}_i^{-1}(\mathbf{p})) dJ_k \\ = \sum_{k=1}^m \int_{J_k} \tilde{\mathbf{f}}^k \cdot \hat{\mathbf{W}}^{ik} dJ_k. \end{aligned}$$

Since the  $J_k$ 's are mutually disjoint and the geometric junction constraints are local to each  $J_k$  individually, the last variational equation can hold if and only if

$$\begin{aligned} (2.11) \quad \sum_{i \in I(J_k)} \int_{J_k} \left( \sum_{\alpha,j} \nu_\alpha^i s_i^{\alpha,j} \mathbf{a}_j^i + \sum_{\alpha,\beta} \nu_\alpha^i s_i^{\alpha,\beta} \mathbf{W}_{,\beta}^i \right) \cdot \hat{\mathbf{W}}^i(\mathbf{p}_i^{-1}(\mathbf{p})) dJ_k \\ = \int_{J_k} \tilde{\mathbf{f}}^k \cdot \hat{\mathbf{W}}^{ik} dJ_k, \quad k = 1, \dots, m. \end{aligned}$$

From (2.3) we have

$$\hat{\mathbf{W}}^i(\mathbf{p}_i^{-1}(\mathbf{p})) = \hat{\mathbf{W}}^{ik}(\mathbf{p}_{ik}^{-1}(\mathbf{p})), \quad \forall i \in I(J_k).$$

Since  $\hat{\mathbf{W}}^{ik}(\mathbf{p}_{ik}^{-1}(\mathbf{p}))$  is arbitrary, it follows that

$$(2.12) \quad \sum_{i \in \mathcal{I}(J_k)} \left( \sum_{\alpha, j} \nu_{\alpha}^i s_i^{\alpha j} \mathbf{a}_j^i + \sum_{\alpha, \beta} \nu_{\alpha}^i s_i^{\alpha \beta} \mathbf{W}_{i, \beta}^i \right) (\mathbf{p}_i^{-1}(\mathbf{p}), t) \\ = \tilde{\mathbf{f}}^k(\mathbf{p}_{ik}^{-1}(\mathbf{p}), t), \quad \mathbf{p} \in J_k.$$

Equation (2.12) is a balance law for forces along the joint  $J_k$  and we refer to it as the *dynamic junction condition* there.

**2.3. Linearization.** The nonlinear equations of motion (2.8), dynamic boundary conditions (2.9) and dynamic junction conditions (2.12) are geometrically exact under the kinematic hypotheses imposed above (i.e., (1.1) and Hooke's Law). It is apparent that their linearizations about the equilibrium  $\mathbf{W}^i = 0$ ,  $i = 1, \dots, n$ , are the same as the linearizations of the equations

$$(2.13) \quad \sum_{\alpha, j} s_{i, \alpha}^{\alpha j} \mathbf{a}_j^i + \mathbf{F}^i = \rho_i \mathbf{W}_{i, \mu}^i, \quad (x_1, x_2) \in \Omega_i, \quad t > 0;$$

$$(2.14) \quad \sum_{\alpha, j} \nu_{\alpha}^i s_i^{\alpha j} \mathbf{a}_j^i = \mathbf{f}^i \quad \text{on } \Gamma_i^D, \quad t > 0;$$

$$(2.15) \quad \sum_{i \in \mathcal{I}(J_k)} \left( \sum_{\alpha, j} \nu_{\alpha}^i s_i^{\alpha j} \mathbf{a}_j^i \right) (\mathbf{p}_i^{-1}(\mathbf{p}), t) = \tilde{\mathbf{f}}^k(\mathbf{p}_{ik}^{-1}(\mathbf{p}), t), \\ \mathbf{p} \in J_k, \quad t > 0, \quad k = 1, \dots, m.$$

The linearizations of (2.13)–(2.15) are obtained by replacing the nonlinear strain components  $\varepsilon_{\alpha, j}$  in the stress-strain law (1.7) by their linear approximations

$$(2.16) \quad \varepsilon_{\alpha, j} \approx \frac{1}{2}(W_{\alpha, j}^i + W_{j, \alpha}^i).$$

Let us consider the case of an elastically isotropic body  $\mathcal{B}_i$ . Its stress-strain law is then

$$s_i^{jk} = 2\mu_i \varepsilon_{jk}^i + \lambda_i \sum_l \varepsilon_{ll}^i \delta_{jk},$$

where  $\varepsilon_{jk}^i$  are the linearized strains in  $\mathcal{B}_i$  and  $\lambda_i$  and  $\mu_i$  are its Lamé parameters. It follows from (2.16) that the linearized stress-displacement relation is

$$(2.17) \quad \begin{cases} s_i^{\alpha j} = \mu_i (W_{\alpha, j}^i + W_{j, \alpha}^i) + \lambda_i \sum_{\gamma} W_{\gamma, \gamma}^i \delta_{\alpha, j}, \\ s_i^{\alpha 3} = \mu_i W_{3, \alpha}^i. \end{cases}$$

With the stresses given by (2.17), the system (2.13) may be written

$$(2.18) \quad \begin{cases} \sum_{\alpha} s_{i,\alpha}^{\alpha\beta} + F_{\beta}^i = \rho_i W_{\beta,tt}^i, \\ \mu_i \Delta W_3^i + F_3^i = \rho_i W_{3,tt}^i \quad \text{in } \Omega, t > 0, \end{cases}$$

where  $F_j^i = \mathbf{F}^i \cdot \mathbf{a}_j^i$  and  $\Delta W_3^i = \sum_{\alpha} W_{3,\alpha\alpha}^i$ . The boundary conditions (2.14) linearize to

$$(2.19) \quad \begin{cases} \sum_{\alpha} \nu_{\alpha}^i s_i^{\alpha\beta} = f_{\beta}^i, \\ \mu_i \boldsymbol{\nu}^i \cdot \nabla W_3^i = f_3^i \quad \text{on } \Gamma_i^D, t > 0, \end{cases}$$

where  $f_j^i = \mathbf{F}^i \cdot \mathbf{a}_j^i$  and  $\nabla W_3^i = \sum_{\alpha} W_{3,\alpha}^i \mathbf{a}_{\alpha}^i$ , and the dynamic junction condition (2.15) becomes

$$(2.20) \quad \sum_{i \in \mathcal{I}(J_k)} \sum_{\alpha} \nu_{\alpha}^i \left( \sum_{\beta} s_i^{\alpha\beta} \mathbf{a}_{\beta}^i + \mu_i W_{3,\alpha}^i \mathbf{a}_3^i \right) (\mathbf{p}_i^{-1}(\mathbf{p}), t) = \tilde{\mathbf{F}}^k(\mathbf{p}_{i_k}^{-1}(\mathbf{p}), t), \quad \mathbf{p} \in J_k, t > 0.$$

The description of the linear system is completed by adjoining to (2.18)–(2.20) the geometric conditions

$$(2.21) \quad \mathbf{W}^i = 0 \quad \text{on } \Gamma_i^C, t > 0,$$

$$(2.22) \quad \mathbf{W}^i(\mathbf{p}_i^{-1}(\mathbf{p}), t) = \mathbf{W}^j(\mathbf{p}_j^{-1}(\mathbf{p}), t), \quad \mathbf{p} \in J_k; i, j \in \mathcal{I}(J_k); k = 1, \dots, m,$$

and the initial conditions

$$(2.23) \quad \mathbf{W}^i|_{t=0} = \mathbf{W}_0^i, \quad \mathbf{W}_t^i|_{t=0} = \mathbf{W}_1^i \quad \text{in } \Omega, i = 1, \dots, n.$$

Equations (2.18)–(2.22), with  $s_i^{jk}$  given by (2.17), describe the linearized motion of a system of interconnected isotropic elastic membranes. Let us note that the above system is also directly derivable from the variational principle

$$\delta \int_0^T [\mathcal{K}(t) - \mathcal{U}(t) + \mathcal{W}(t)] dt = 0,$$

where  $\mathcal{K}$  and  $\mathcal{W}$  are as above and

$$\mathcal{U}(t) = \frac{1}{2} \sum_{i=1}^n \int_{\Omega_i} \left( \sum_{\alpha, \beta} s_i^{\alpha\beta} W_{\beta, \alpha}^i + \mu_i |\nabla W_3^i|^2 \right) d\Omega.$$



### 3. Controllability of Linked Isotropic Membranes

In this section we shall consider the question of exact controllability of solutions of the linearized equations of motion of the system of interconnected, elastically isotropic membranes given by (2.18)–(2.23). Because of space limitations, only an overview of this problem can be presented.

It is assumed that all distributed forces vanish. Given a suitable  $T > 0$  and "arbitrary" initial data  $(\mathbf{W}_0^i, \mathbf{W}_1^i)$  and final data  $(\hat{\mathbf{W}}_0^i, \hat{\mathbf{W}}_1^i)$ , the above system is *exactly controllable in time  $T$*  if there are control functions  $\mathbf{f}^i = \sum_j f_j^i \mathbf{a}_j^i$ ,  $\mathbf{f}^k$  such that the solution of (2.18)–(2.23) achieves the state  $(\hat{\mathbf{W}}_0^i, \hat{\mathbf{W}}_1^i)$  at time  $T$ :

$$\mathbf{W}^i|_{t=T} = \hat{\mathbf{W}}_0^i, \quad \mathbf{W}_t^i|_{t=T} = \hat{\mathbf{W}}_1^i, \quad i = 1, \dots, n.$$

Of course, the data and controls must be chosen from appropriate function spaces for which the initial value problem is (at least) well-posed in some sense. Since the problem (2.18)–(2.22) is time reversible, one may assume that either the final state or the initial state vanishes. In the latter situation the exact controllability problem is called the *reachability problem*. Thus the reachability problem is that of showing that starting from the zero state, an "arbitrary state"  $(\hat{\mathbf{W}}_0^i, \hat{\mathbf{W}}_1^i)$ ,  $i = 1, \dots, n$ , may be achieved in time  $T$  through an appropriate choice of control functions.

It is not to be expected that every system of interconnected membranes, no matter how configured, is exactly controllable. Indeed, our purpose is to determine those geometric properties of such a system which will guarantee its exact controllability. It will be fairly apparent that the sufficient conditions for exact controllability discussed below are rather far from necessary.

Another point is that in order to obtain any exact controllability results at all we usually (but not always) find it necessary to employ controls not only along the exterior boundaries  $L_i^D$  but also in the junction regions  $L_i^J$ . It is not clear to what extent the latter requirement is due to the methods employed or whether there is something intrinsic to wave propagation in general systems of interconnected membranes which necessitates the use of junction based controls.

**3.1. Observability Estimates for the Homogeneous Problem.** Whether one proceeds via the control-to-state map or employs the (equivalent) methodology of the Hilbert Uniqueness Method, it is well-known that the reachability problem is the "dual" of the *continuous observability problem* for the homogeneous adjoint system (which in the present case coincides with the original system). When dealing with situations in which the controls act on the boundary, the latter problem involves showing that certain traces of the solution on the boundary may be estimated from below by some norm of the initial data. Each such *observability estimate* leads to a reachability result, and conversely. Here we shall describe three such observability estimates.

Let us therefore consider the homogeneous problem describing the motion of a system of interconnected membranes. Let the displacement vector of the  $i$ -th

membrane be

$$(3.1) \quad \Psi^i = \sum_j \Psi_j^i \mathbf{a}_j^i = \boldsymbol{\psi}^i + \Psi_3^i \mathbf{a}_3^i, \quad \boldsymbol{\psi}^i = \sum_\alpha \Psi_\alpha^i \mathbf{a}_\alpha^i.$$

We write

$$\begin{aligned} \varepsilon_{\alpha\beta}(\boldsymbol{\psi}^i) &= \frac{1}{2} (\Psi_{\alpha,\beta}^i + \Psi_{\beta,\alpha}^i), \\ s_i^{\alpha\beta}(\boldsymbol{\psi}^i) &= 2\mu_i \varepsilon_{\alpha\beta}(\boldsymbol{\psi}^i) + \lambda_i \sum_\gamma \varepsilon_{\gamma\gamma}(\boldsymbol{\psi}^i) \delta_{\alpha\beta}. \end{aligned}$$

The system under consideration is then

$$(3.2) \quad \begin{cases} \sum_\alpha s_{i,\alpha}^{\alpha\beta}(\boldsymbol{\psi}^i) = \rho_i \Psi_{3,t}^i, \\ \mu_i \Delta \Psi_3^i = \rho_i \Psi_{3,t}^i, \quad \mathbf{x} \in \Omega_i, \quad t > 0; \end{cases}$$

$$(3.3) \quad \begin{cases} \sum_\alpha \nu_\alpha^i s_i^{\alpha\beta}(\boldsymbol{\psi}^i) = 0, \\ \boldsymbol{\nu}^i \cdot \nabla \Psi_3^i = 0 \quad \mathbf{x} \in \Gamma_i^D, \quad t > 0; \end{cases}$$

$$(3.4) \quad \Psi^i = 0, \quad \mathbf{x} \in \Gamma_i^C, \quad t > 0;$$

$$(3.5) \quad \Psi^i(\mathbf{p}_i^{-1}(\mathbf{p}), t) = \Psi^j(\mathbf{p}_j^{-1}(\mathbf{p}), t), \quad \mathbf{p} \in J_k, \quad i, j \in \mathcal{I}(J_k), \quad k = 1, \dots, m;$$

$$(3.6) \quad \sum_{i \in \mathcal{I}(J_k)} \sum_\alpha \nu_\alpha^i \left( \sum_{\beta} s_i^{\alpha\beta}(\boldsymbol{\psi}^i) \mathbf{a}_\beta^i + \mu_i \Psi_{3,\alpha}^i \mathbf{a}_3^i \right) (\mathbf{p}_i^{-1}(\mathbf{p}), t) = 0, \quad \mathbf{p} \in J_k, \quad t > 0;$$

$$(3.7) \quad \Psi^i|_{t=0} = \Psi_0^i, \quad \Psi_{3,t}^i|_{t=0} = \Psi_1^i, \quad \mathbf{x} \in \Omega_i.$$

We set  $\boldsymbol{\Psi}_0 = (\Psi_0^1, \dots, \Psi_0^n)$  and  $\boldsymbol{\Psi}_1 = (\Psi_1^1, \dots, \Psi_1^n)$ .

Let  $\boldsymbol{\Psi} = (\boldsymbol{\Psi}^1, \dots, \boldsymbol{\Psi}^n)$  with  $\boldsymbol{\Psi}^i : \Omega_i \mapsto \mathbb{R}^3$  given by (3.1). Let  $H$  (resp.,  $V$ ) be the Hilbert space consisting of those  $\boldsymbol{\Psi}$  for which  $\Psi_j^i \in L^2(\Omega_i)$  (resp.,  $\Psi_j^i \in H^1(\Omega_i)$ ), with the norms of  $\boldsymbol{\Psi}$  in  $H$  and in  $V$  given by

$$\|\boldsymbol{\Psi}\|_H = \left( \sum_{i=1}^n \sum_j \|\Psi_j^i\|_{L^2(\Omega_i)}^2 \right)^{1/2}$$

and

$$\|\boldsymbol{\Psi}\|_V = \left( \sum_{i=1}^n \sum_j \|\Psi_j^i\|_{H^1(\Omega_i)}^2 \right)^{1/2},$$

respectively. Let  $\mathcal{V}$  denote the closed subspace of  $V$  consisting of those  $\Psi \in V$  satisfying

$$\begin{aligned}\Psi^i &= 0 \quad \text{on } \Gamma_i^C, \\ \Psi^i(\mathbf{p}_i^{-1}(\mathbf{p})) &= \Psi^j(\mathbf{p}_j^{-1}(\mathbf{p})), \quad \mathbf{p} \in J_k, \quad i, j \in \mathcal{I}(J_k).\end{aligned}$$

The space  $\mathcal{V}$  is dense in  $H$  with compact embedding.

Let  $\mathbf{q}_0$  be a point of  $\mathbb{R}^3$  and consider the following restrictions on the geometry of the configuration of membranes and on the material parameters:

$$(3.8) \quad (\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^i \leq 0 \quad \text{on } \Gamma_i^C;$$

$$(3.9) \quad \sum_{i \in \mathcal{I}(J_k)} \rho_i (\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^i \leq 0 \quad \text{on } J_k, \quad k = 1, \dots, m;$$

$$(3.10) \quad \sum_{i \in \mathcal{I}(J_k)} a_i = 0, \quad \sum_{i \in \mathcal{I}(J_k)} a_i^2 = 1 \implies \sum_{i \in \mathcal{I}(J_k)} \frac{a_i^2}{\mu_i} (\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^i < 0 \quad \text{on } J_k, \quad k = 1, \dots, m;$$

and

$$(3.11) \quad \sum_{i \in \mathcal{I}(J_k)} (b_i \boldsymbol{\nu}^i + c_i \mathbf{a}_3^i) = 0, \quad \sum_{i \in \mathcal{I}(J_k)} (b_i^2 + c_i^2) = 1 \implies \sum_{i \in \mathcal{I}(J_k)} \left( \frac{b_i^2}{2\mu_i + \lambda_i} + \frac{c_i^2}{\mu_i} \right) (\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^i < 0 \quad \text{on } J_k, \quad k = 1, \dots, m.$$

Further, let us introduce the sets

$$\Lambda_i^{D^+} = \{\mathbf{p} \in \Lambda_i^D \mid (\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^i \geq 0\}, \quad \Lambda_i^D = \Lambda_i^{D^+} / \Lambda_i^D,$$

$$\Gamma_i^{D^+} = \mathbf{p}_i^{-1}(\Lambda_i^{D^+}), \quad \Gamma_i^D = \mathbf{p}_i^{-1}(\Lambda_i^D) = \Gamma_i^D / \Gamma_i^{D^+}.$$

We have the following observability estimate.

**Theorem 3.1.** *Assume that  $\mathbf{q}_0$  may be chosen so that (3.8)–(3.11) hold and further suppose that either*

- (i)  $\Gamma_i^{C^+} \neq \emptyset$  for at least one index  $i$ , or
- (ii)  $\Gamma_i^{D^+} \neq \emptyset$  for at least one index  $i$ .

If  $\Psi = (\Psi^1, \dots, \Psi^n)$  is a sufficiently regular solution of (3.2)–(3.7), then there is a  $T_0 > 0$  such that for  $T > T_0$  there is a constant  $C_T$  such that

$$(3.12) \quad \|(\Psi_0, \Psi_1)\|_{V \times H}^2 \leq C_T \left\{ \sum_{i=1}^n \left[ \int_0^T \int_{\Gamma_i^D} (|\Psi_{,t}^i|^2 + |\Psi^i|^2) d\Gamma dt \right. \right. \\ \left. \left. + \int_0^T \int_{\Gamma_i^D} \left| \frac{\partial \Psi^i}{\partial \tau^i} \right|^2 d\Gamma dt \right] + \sum_{k=1}^m \int_0^T \int_{\Gamma_{i_k}(J_k)} \left| \frac{\partial \Psi^{i_k}}{\partial \tau^{i_k}} \right|^2 d\Gamma dt \right\}.$$

**Remark 3.1.** The problem (3.2)–(3.7) has a natural variational formulation (the principle of virtual work) from which existence and uniqueness of a solution may be proved. The variational solution has components  $\Psi_j^i$  which belong to  $H^1(\Omega_i)$  and the  $\Psi^i$  satisfy the geometric junction and boundary conditions, provided the initial data do likewise. In order to prove Theorem 3.1, however, we need to require that the solution have additional regularity, namely that  $\Psi_j^i \in H^s(\Omega_i)$  for some  $s > 3/2$  provided that the initial data are sufficiently regular. It is known that even for a single isotropic membrane this degree of regularity depends on the geometry of  $\Omega$  and the particular boundary conditions. For example, it is true for any Lipschitz domain if  $\bar{\Gamma}^C$  and  $\bar{\Gamma}^D$  are disjoint. If  $\bar{\Gamma}^C \cap \bar{\Gamma}^D \neq \emptyset$ , the desired regularity holds if these sets meet in an angle smaller than  $\pi$  (measured in the interior of  $\Omega$ ). In this matter the reader is referred to Grisvard [1], [2], and Nicaise [4]. For systems of interconnected isotropic membranes, from the regularity results for a single membrane it is clear that there is  $H^s$  regularity ( $s > 3/2$ ) in a neighborhood of each point of  $\Gamma_i^C$ , of  $\Gamma_i^D$  and of  $\bar{\Gamma}_i^C \cap \bar{\Gamma}_i^D$  provided the angle condition just mentioned is satisfied. (Interior regularity is not at issue.) However, the precise regularity possessed by  $\Psi^i$  near  $\Gamma_i^J$  or, more specifically, its regularity near points of  $\bar{\Gamma}_i^J \cap \bar{\Gamma}_i^C$  and  $\bar{\Gamma}_i^J \cap \bar{\Gamma}_i^D$  is, to the knowledge of this author, unsettled and requires further analysis. Let us mention, however, an important contribution of Nicaise [7] towards resolving this issue, wherein the precise singular behavior of solutions of the Laplace and biharmonic equations on networks at mixed corners is described. See also Nicaise [5], [6], where regularity of solutions (among other things) of "transmission problems" for general elliptic problems with general boundary and interface conditions is investigated.

**Remark 3.2.** A sufficient condition for the validity of (3.8)–(3.11) is that, at each joint  $J_k$ , at most one of the  $(\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^i$ , say  $(\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^{q_k}$ , is positive,  $(\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^i < 0$  if  $i \neq q_k$ ,  $i \in \mathcal{I}(J_k)$ , and

$$\rho_{q_k}(\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^{q_k}, \quad (\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^{q_k} / \mu_{q_k}, \quad (\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^{q_k} / \lambda_{q_k}$$

are sufficiently small relative to

$$\rho_i |(\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^i|, \quad |(\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^i| / \mu_i, \quad |(\mathbf{p} - \mathbf{q}_0) \cdot \boldsymbol{\nu}^i| / \lambda_i,$$

respectively, for all  $i \neq q_k$  and  $i \in \mathcal{I}(J_k)$ . This means that  $\rho_{q_k}$  has to be sufficiently small relative to the other  $\rho_i$ 's and that  $\mu_{q_k}, \lambda_{q_k}$  must be sufficiently large relative to the other  $\mu_i$ 's and  $\lambda_i$ 's, respectively.

The observability estimate (3.12) leads to a corresponding reachability result (Theorem 3.3 below). Because of the presence of integrals over  $J_k$  in (3.12), this reachability result requires controls in the junction regions as well as on the exterior boundaries  $\Gamma_i^D$ . We shall now discuss one particular configuration of connected membranes (there are others as well) for which the integrals over  $J_k$  may be eliminated from (3.12), namely the one in which the membranes are *serially connected*. This means that

$$\begin{cases} \mathbf{a}_3^i = \pm \mathbf{a}_3^j, & \forall i, j = 1, \dots, n, \\ m = n - 1. \end{cases}$$

By reindexing the membranes and joints, if necessary, we may assume that

$$\begin{cases} \overline{\mathcal{P}}_i \cap \overline{\mathcal{P}}_{i+1} = J_i, & i = 1, \dots, n - 1, \\ \overline{\mathcal{P}}_i \cap \overline{\mathcal{P}}_j = \emptyset, & |i - j| > 1. \end{cases}$$

Thus  $\mathcal{I}(J_k) = [k, k + 1]$ . Further, without loss of generality, we may suppose that  $\mathbf{a}_3^i = \mathbf{a}_3^j = (0, 0, 1)$  (hence

$$\boldsymbol{\nu}^i = -\boldsymbol{\nu}^{i+1}, \quad \boldsymbol{\tau}^i = -\boldsymbol{\tau}^{i+1} \quad \text{on } J_i)$$

and that

$$\begin{cases} \mathcal{P}_i = \Omega_i, \quad \Lambda_i^C = \Gamma_i^C, \quad \Lambda_i^D = \Gamma_i^D, \\ \Gamma_k(J_k) = \Gamma_{k+1}(J_k) = J_k, \quad k = 1, \dots, n - 1, \\ \Gamma_i(J_k) = \emptyset, \quad |i - k| > 1. \end{cases}$$

If  $\mathcal{P}_i$  is serially connected to  $\mathcal{P}_{i+1}$ , the geometric and dynamic conditions along  $J_i$  reduce to, respectively,

$$(3.13) \quad \boldsymbol{\psi}^i = \boldsymbol{\psi}^{i+1}, \quad \Psi_3^i = \Psi_3^{i+1} \quad \text{along } J_i,$$

and

$$(3.14) \quad \begin{cases} \sum_{\alpha, \beta} \nu_\alpha^{i+1} s_{i+1}^{\alpha\beta} (\boldsymbol{\psi}^{i+1}) \mathbf{a}_{\beta}^{i+1} = - \sum_{\alpha, \beta} \nu_\alpha^i s_i^{\alpha\beta} (\boldsymbol{\psi}^i) \mathbf{a}_{\beta}^i, \\ \mu_i \frac{\partial \Psi_3^i}{\partial \boldsymbol{\nu}^i} = -\mu_{i+1} \frac{\partial \Psi_3^{i+1}}{\partial \boldsymbol{\nu}^{i+1}} \quad \text{along } J_i, \end{cases}$$

so that the transverse motion is not coupled to the in-plane motion, as is to be expected.

When membranes are serially connected, it is possible to eliminate the integrals over  $J_k$  in the estimate (3.12), provided the parameters  $\rho_i, \mu_i, \lambda_i$  satisfy certain monotonicity conditions.

**Theorem 3.2.** Assume that  $\mathbf{x}_0 \in \mathbb{R}^2$  may be chosen so that

$$(3.15) \quad (\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^i = 0 \text{ on } \Gamma_i^i, \quad i = 1, \dots, n.$$

Suppose further that for the given choice of  $\mathbf{x}_0$  and each  $i = 1, \dots, n-1$ ,

$$(3.16) \quad \rho_i \leq \rho_{i+1}, \quad \mu_i \geq \mu_{i+1}, \quad \lambda_i \geq \lambda_{i+1}$$

if  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^i > 0$  on  $J_i$ ;

$$(3.17) \quad \rho_i \geq \rho_{i+1}, \quad \mu_i \leq \mu_{i+1}, \quad \lambda_i \leq \lambda_{i+1}$$

if  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^i < 0$  on  $J_i$ . If  $\Psi = (\Psi^1, \dots, \Psi^n)$  is a sufficiently regular solution of (3.2)–(3.4), (3.13), (3.14) with initial data  $(\Psi_0, \Psi_1)$ , there is a  $T_0 > 0$  such that for each  $T > T_0$  there is a constant  $C_T$  such that

$$(3.18) \quad \|(\Psi_0, \Psi_1)\|_{\mathcal{V} \times H}^2 \leq C_T \sum_{i=1}^n \left\{ \int_0^T \int_{\Gamma_i^{D,i}} (|\Psi_{,t}^i|^2 + |\Psi^i|^2) d\Gamma dt + \int_0^T \int_{\Gamma_i^D} \left| \frac{\partial \Psi^i}{\partial \boldsymbol{\tau}^i} \right|^2 d\Gamma dt \right\}.$$

**Remark 3.3.** Either condition (i) or condition (ii) of the previous theorem will automatically be satisfied in the serial case. The quantity  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^i$  must be of constant sign on  $J_i$ . If, for a particular index  $i$ ,  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^i = 0$  on  $J_i$ , no relation between the material parameters of  $\mathcal{P}_i$  and  $\mathcal{P}_{i+1}$  need be assumed. Also, since the choice of an  $\mathbf{x}_0$  satisfying (3.15) is not unique, for certain configurations it may be possible to choose distinct values of  $\mathbf{x}_0$  such that for certain indices  $i$ ,  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^i > 0$  on  $J_i$  for the first choice of  $\mathbf{x}_0$  while, for the second,  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^i < 0$  on  $J_i$ . In such situations either (3.16) or (3.17) is sufficient for the conclusion of the theorem. These observations suggest that the above restrictions on the material parameters are due to the method of proof employed and are not intrinsic to the validity of the theorem.

**Remark 3.4.** Theorem 3.2 is valid also in the more general situation where the  $J_i$ 's are any Lipschitz continuous curves. Of course, in this situation  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^i$  need not be of constant sign on  $J_i$  so that it is necessary to assume *both*  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^i \geq 0$  on  $J_i$  and (3.16), or  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^i \leq 0$  on  $J_i$  and (3.17). In particular, Theorem 3.2 is valid for the following *problem of transmission*. Let  $\Omega, \Omega_1$  be bounded, open, connected sets in  $\mathbb{R}^2$  with Lipschitz continuous boundaries and with  $\bar{\Omega}_1 \subset \Omega$ , and set  $\Omega_2 = \Omega/\bar{\Omega}_1$ . We consider the above problem with  $J = \partial\Omega_1$  being the junction region. We then have  $\partial\Omega = \bar{\Gamma}_2^C \cup \bar{\Gamma}_2^D$ . Suppose that there is a point  $\mathbf{x}_0 \in \mathbb{R}^2$  such that  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^1 > 0$  on  $J$ ,  $(\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\nu}^2 \leq 0$  on  $\Gamma_2^C$  and assume that (3.16) holds. Then the estimate (3.18) is valid. Since, for serial membranes, there is no coupling between in-plane and transverse motions, (3.18) comprises two separate observability estimates, one for transverse motion alone (by setting  $\psi^i = 0$ ) and

one for in-plane motion alone (by setting  $\Psi_4^i = 0$ ). The function  $\Psi_4^i$  is a solution of the problem

$$(3.19) \quad \Psi_{4,i}^i - \mu_i \Delta \Psi_4^i = 0 \text{ in } \Omega, \quad i = 1, 2.$$

$$(3.20) \quad \Psi_3^1 = \Psi_3^2, \quad \mu_1 \frac{\partial \Psi_3^1}{\partial \nu^1} = -\mu_2 \frac{\partial \Psi_3^2}{\partial \nu^2} \text{ on } J,$$

$$(3.21) \quad \Psi_3^2 = 0 \text{ on } \Gamma_2^U, \quad \partial \Psi_3^2 / \partial \nu^2 = 0 \text{ on } \Gamma_2^D.$$

An observability estimate analogous to (3.18) for the problem consisting of (3.19), (3.20) and the Dirichlet boundary condition  $\Psi_3^2 = 0$  on  $\partial\Omega$  in place of (3.21), has been proved by Lions [3, Theorem VI.4.2].

The following corollary is an immediate consequence of Theorem 3.2 and the process of "weakening the norm" (see, e.g., [3, p. 200]). Let  $\mathcal{V}'$  denote the dual space of  $\mathcal{V}$  with respect to  $H$ .

**Corollary 3.1.** *In addition to the assumptions of Theorem 3.2, suppose that*

$$(\mathbf{x} - \mathbf{x}_0) \cdot \nu^i \geq 0 \text{ on } \Gamma_i^D$$

(i.e.,  $\Gamma_i^D = \emptyset$ ). Then there is a  $T_0 > 0$  such that for each  $T > T_0$  there is a constant  $C_T$  such that

$$\|(\Psi_0, \Psi_1)\|_{H \times \mathcal{V}'}^2 \leq C_T \sum_{i=1}^2 \int_0^T \int_{\Gamma_i^D} |\Psi_{3,i}^i|^2 d\Gamma dt.$$

**3.2. The Reachable States.** Each of the above observability estimates leads to a reachability result. Here we indicate two such results, one being the "dual" of Theorem 3.1 and the other the dual of Corollary 3.1. Since the observability estimates are known only for "sufficiently regular" solutions, implicit in the theorems below is the assumption that there is a set of initial data which is dense in  $\mathcal{V} \times H$  and for which the corresponding solution of (3.2)–(3.7) has the desired regularity. This assumption imposes a further constraint on the configuration of the system of membranes, but we do not know how to quantify this constraint in terms of simple geometric properties of the configuration. (See Remark 3.1 above).

**Theorem 3.3.** *Let the hypotheses of Theorem 3.1 hold and let  $T > T_0$  and assume that  $\mathbf{W}_0 \in H$ ,  $\mathbf{W}_1 \in \mathcal{V}'$ . Then there are controls  $\mathbf{f}^i$ ,  $\tilde{\mathbf{f}}^k$  such that the solution of (2.18)–(2.22) with  $\mathbf{W}|_{t=0} = \mathbf{W}_0$ ,  $\mathbf{W}_{,t}|_{t=0} = 0$  satisfies*

$$\mathbf{W}|_{t=T} = \mathbf{W}_0, \quad \mathbf{W}_{,t}|_{t=T} = \mathbf{W}_1.$$

The controls of Theorem 3.3 belong to the following (very weak) spaces:

$$\mathbf{f}' \in (H^1(\Omega, T; \mathcal{L}^2(\Gamma_i^{D'})))' \bigoplus L^2(0, T; (\mathcal{H}^1(\Gamma_i^{D'})))',$$

$$\hat{\mathbf{f}}^k \in L^2(0, T; (\mathcal{H}^1(\Gamma_k(J_k))))',$$

the primes denoting dual spaces, where

$$\mathcal{L}^2(\Gamma_i^{D'}) = \left\{ \sum_j f_j \mathbf{a}_j \mid f_j \in L^2(\Gamma_i), \text{ supp } f_j \subset \Gamma_i^{D'} \right\}$$

and

$$\mathcal{H}^1(\Gamma_i^{D'}) = \left\{ \hat{\Psi} \in \mathcal{L}^2(\Gamma_i^{D'}) \mid \frac{\partial \hat{\Psi}}{\partial \boldsymbol{\tau}^i} \in \mathcal{L}^2(\Gamma_i^{D'}) \right\}.$$

The corresponding solution of (2.18)–(2.22) with  $\mathbf{W}|_{t=0} = \mathbf{W}_0|_{t=0} = 0$  exists in a very weak sense which may be made precise through the method of transposition, for example.

In the case of serially connected membranes, a more satisfactory result follows from Corollary 3.1.

**Theorem 3.4.** *Let the hypotheses of Corollary 3.1 hold. Then*

$$\mathcal{V} \times H \subset \{(\mathbf{W}(T), \mathbf{W}_0(T)) \mid \mathbf{f}' \in L^2(0, T; \mathcal{L}^2(\Gamma_i^{D'})) \text{ and } \hat{\mathbf{f}}^k = 0\}.$$

Thus, for serially connected membranes, all finite energy states are reachable using  $L^2(\Gamma_i^{D'} \times (0, T))$  controls only.

## References

1. P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman, London, 1985.
2. ———, "Singularités en élasticité", *J. Math. Pures Appl.* **107** (1989), 157–180.
3. J. L. Lions, *Contrôlabilité Exacte, Perturbations et Stabilisation de Systèmes Distribués, Tome I: Contrôlabilité Exacte*, Collection RMA, vol. 8, Masson, Paris, 1988.
4. S. Nicaise, "About the Lamé system in a polygonal or a polyhedral domain and a coupled problem between the Lamé system and the plate equation. I: regularity of the solutions", *Ann. Scuola Normale Sup. Pisa* **XIX** (1992), 327–361.
5. ———, "General interface problems I", *Math. Methods Appl. Sci.*, to appear.
6. ———, "General interface problems II", *Math. Methods Appl. Sci.*, to appear.
7. ———, *Polygonal Interface Problems*, Peter Lang, Frankfurt am Main, 1993.
8. G. Wempner, *Mechanics of Solids with Applications to Thin Bodies*, Sijthoff and Noordhoff, Rockville, Maryland, 1981.

J. E. Lagnese  
Department of Mathematics  
Georgetown University  
Washington, DC 20057, USA



## ON FEEDBACK CONTROLS FOR DYNAMIC NETWORKS OF STRINGS AND BEAMS AND THEIR NUMERICAL SIMULATION

GÜNTER LEUGERING

Mathematisches Institut  
Universität Bayreuth

**ABSTRACT.** In these notes we want to present some control strategies for dynamic networks of strings and beams in those situations where the classical concepts of exact or approximate controllability fail. This is, in particular, the case for networks containing circuits. Generically, a residual motion settles in such circuits, even if all nodes are subject to controls. In those situations we resort to controls which direct the flux of energy. Using such controls in a network, we are able to steer the entire energy to preassigned parts of the structure. In practice these parts are more massive and can absorb energy more easily than the fragile elements. We also provide numerical evidence for the control strategies discussed in these notes. The material is related to joint work with J.E. Lagnese and E.G.P.G. Schmidt, and is essentially included [11].

*1991 Mathematics Subject Classification.* 39C20, 39B52, 65M06, 65M60

*Key words and phrases.* Absorbing, directing and nonlinear controls, disappearing solutions, numerical simulation.

### 1. Introduction

In this note we want to give some examples of feedback control strategies applied to dynamic 1-d-networks of strings and Timoshenko beams. We will consider examples of scalar 'out-of-the-plane' displacement models as well as planar systems. Systems of the type considered here have been introduced in Schmidt [19], Leugering and Schmidt [15] and Lagnese, Leugering and Schmidt [13], [14]. It turns out that, speaking in general terms, tree-like structures of strings and Timoshenko beams are exactly controllable in finite time (the time being related to the longest path in the tree), if the root (any one of the simple nodes to be introduced below) is clamped and all of the other simple vertices are under control. As for Euler-Bernoulli beams it was shown in that a star-like bundle of beams is exactly controllable in finite time, if all simple vertices are being controlled. On the negative side, it is known for quite a while, see [4], that already for two strings (scalar displacement) controllability/stabilizability from multiple nodes is sensitive to the

choice of boundary conditions for the simple (uncontrolled) nodes. A similar phenomenon is also observed for tree-like networks. Take for instance the case of three strings coupled at one node: if two simple nodes are clamped and the remaining simple node is controlled, not even approximate controllability can be achieved. The situation, however, is much improved if one of the previously clamped simple nodes is released (satisfies Neumann boundary conditions). Then it can be shown that spectral controllability holds. The picture becomes even more pessimistic if the network contains circuits. In this case, approximate controllability can not be expected, even if all nodes are controlled. As real multi-link flexible structures are, indeed, almost exclusively repetitive, the classical control concepts will usually fail. Even if, from a mathematical point of view, properties like 'rational independence of optical lengths' can yield sharper controllability results, those results are only of academic interest, as it is impossible to produce those networks in reality. In addition, boundary conditions are usually subject to various disturbances, so that a control strategy based on the classical concepts might work in one situation and fail in another. The content of these notes is essentially contained in the above mentioned joint work with J.E. Lagnese and E.J.P.G. Schmidt, [14]. The purpose is to introduce control strategies which do not aim at complete absorption in finite time (controllability) or global decay of the total energy (stability) but, rather, make it possible to channel the energy through the network to end up at a set of specified 'safe' nodes where energy is more easily absorbed (for instance at the base of the structure). The major emphasis is on numerical evidence for these concepts, rather than on the formulation of theorems. We will, however, provide the arguments for exemplaric cases.

## 2. Notation

We consider a nonempty, finite, simple and connected graph  $\mathbf{G}$  in  $\mathbf{R}^2$  with  $n$  vertices,  $V(\mathbf{G}) := \{v_i, i = 1, \dots, n\}$ , and  $N$  edges  $E(\mathbf{G}) := \{\mathbf{k}_j, j = 1, \dots, N\}$ . The edges  $\mathbf{k}_j$  are parametrized by  $\pi_j : [0, \ell_j] \rightarrow \mathbf{R}^2$ , where the running variable  $x_j \in [0, \ell_j]$  represents the arclength. The maps  $\pi_j$  are assumed to be  $C^2$ -smooth. For  $i, j \in V(\mathbf{G})$  we set  $\mathbf{k}_{s(i,h)} := v_i v_j = v_j v_i$ , where  $v_i v_j$  signifies the edge joining the vertices  $v_i, v_j$ , if  $v_i v_j \in E(\mathbf{G})$ . In fact, as we will be concerned here only with networks consisting of straight edges, we assign to each edge  $\mathbf{k}_{s(i,h)}$  the unit edge vector  $\mathbf{e}_{s(i,h)}$  directed along the edges, and its orthogonal complement  $\mathbf{e}_{s(i,h)}^\perp$ . We also introduce the unit vector  $\mathbf{n}$  normal to the plane of the graph. To each triad  $\mathbf{e}_{s(i,h)}, \mathbf{e}_{s(i,h)}^\perp, \mathbf{n}$  we associate the corresponding set of orthogonal projections  $\mathbf{P}_{s(i,h)}, \mathbf{P}_{s(i,h)}^\perp, \mathbf{P}_n$ . Furthermore, we introduce the incidence matrix

$$(2.1) \quad d_{ij} = \begin{cases} 1 & \text{if } \pi_j(\ell_j) = v_i \\ -1 & \text{if } \pi_j(0) = v_i \\ 0 & \text{else} \end{cases}$$

and the adjacency matrix

$$(2.2) \quad c_{ih} = \begin{cases} 1 & \text{if } v_i v_h \in E(\mathbf{G}) \\ 0 & \text{else} \end{cases}.$$

For each vertex we define the set  $\Gamma(v_i) := \{v_h \in V(\mathbf{G}) | c_{ih} = 1\}$  of all vertices adjacent to the vertex  $v_i$ , and  $d(v_i) := |\Gamma(v_i)|$  the edge degree of the vertex  $v_i$ . We can now characterize the network as follows

$$\begin{aligned} \mathbf{G} &= \bigcup_{j=1}^N \mathbf{k}_j \\ \partial V(\mathbf{G}) &= \{v_i \in V(\mathbf{G}) | d(v_i) = 1\}, \\ V(\mathbf{G})^\circ &= \{v_i \in V(\mathbf{G}) | d(v_i) > 1\}. \end{aligned}$$

Obviously,  $\partial V(\mathbf{G})$  signifies the set of simple nodes, where only one edge starts or ends, while  $V(\mathbf{G})^\circ$  signifies the set of multiple nodes, where more than 1 edge meet. A function  $\mathbf{r} : \mathbf{G} \rightarrow \mathbf{R}^l$  can then be viewed as a collection of functions with values along the individual edges by

$$\mathbf{r}_j := \mathbf{r}(\pi_j(x)), \quad x \in [0, l_j].$$

### 3. Networks of strings

In this section we consider a network of strings which, in its reference configuration, coincides with a graph  $\mathbf{G}$  as described in the first subsection. Let  $\mathbf{r}$  denote the deviation from the reference configuration. By the definition above,  $\mathbf{r}_j$  denotes the displacement of the string (edge) with label  $j$ . We define

$$\begin{aligned} u_j \mathbf{e}_j &:= \langle \mathbf{r}_j, \mathbf{e}_j \rangle \mathbf{e}_j = \mathbf{P}_j \mathbf{r}_j, \\ w_j \mathbf{e}_j^\perp &:= \langle \mathbf{r}_j, \mathbf{e}_j^\perp \rangle \mathbf{e}_j^\perp = \mathbf{P}_j^\perp \mathbf{r}_j, \end{aligned}$$

so that  $u_j, w_j$  represent the longitudinal and vertical displacement of the  $j$ -th string, respectively. Let  $p_j^2, q_j^2$  denote the longitudinal and flexural rigidity. Furthermore, let  $\mathbf{K}_j := p_j^2 \mathbf{P}_j + q_j^2 \mathbf{P}_j^\perp$  be the stiffness operator associated with  $j$ -th string. Then the equations governing the motion of the network of strings can be written down as follows:

$$\begin{aligned} \ddot{\mathbf{r}}_j - \mathbf{K}_j \mathbf{r}_j'' &= 0, \quad j = 1, \dots, N, \\ \mathbf{r}_{s(i,h)}(v_i) &= 0, \quad \forall v_i \in \partial_0 V(\mathbf{G}), \\ \mathbf{K}_{s(i,h)} \mathbf{r}_{s(i,h)}'(v_i) &= \mathbf{f}_{v_i}, \quad \forall v_i \in \partial_1 V(\mathbf{G}), \\ \mathbf{r}_{s(i,h)}(v_i) &= \mathbf{r}_{s(i,j)}(v_i), \quad v_h, v_j \in \Gamma(v_i), \quad v_i \in V(\mathbf{G})^\circ, \\ \sum_{h: v_h \in \Gamma(v_i)} d_{is(i,h)} \mathbf{K}_{s(i,h)} \mathbf{r}_{s(i,h)}'(\pi_{s(i,h)}^{-1}(v_i)) &= \mathbf{g}_{v_i}, \quad \forall v_i \in V(\mathbf{G})^\circ, \\ \mathbf{r}_j(\cdot, 0) &= \mathbf{r}_{j0}, \quad \dot{\mathbf{r}}_j(\cdot, 0) = \dot{\mathbf{r}}_{j1}. \end{aligned}$$

Here  $\partial_0 V(\mathbf{G}) \subset \partial V(\mathbf{G})$  denotes the set of simple nodes with clamped boundary conditions, and  $\partial_1 V(\mathbf{G})$  is the set of stress free (or externally loaded as in the case where  $\mathbf{f}_v \neq 0$ ) simple nodes. The third equation expresses the continuity of the network, or its connectedness also in the deformed configuration, while the fourth equation represents the balance of forces (with a possible external loading modeled through the functions  $\mathbf{g}_v$ ).

#### 4. Networks of Timoshenko beams

The situation is notationally very similar but mathematically more involved in the case where each edge is considered as a Timoshenko beam in its reference configuration. In addition to the longitudinal and vertical displacement  $(u_j, w_j)$  of the centerline of the beam we also have a rotation  $(\psi_j := \langle \mathbf{P}_n \mathbf{r}_j, \mathbf{n} \rangle)$  of the cross section out of its position perpendicular to the centerline, i.e. we have to account for shearing. We denote the shear moduli by  $q_j^2$  and the flexural rigidity by  $m_j^2$ , as before  $p_j^2$  denotes the longitudinal stiffness. We only consider here an initially straight and untwisted linear Timoshenko-Bresse system as described above.

$$(4.1) \quad \begin{aligned} \ddot{u}_j - p_j^2 u_j'' &= 0, \\ \ddot{w}_j - q_j^2 (w_j' + \psi_j)' &= 0, \\ \ddot{\psi}_j - m_j^2 \psi_j'' + q_j^2 (w_j' + \psi_j) &= 0, \end{aligned}$$

$$(4.2) \quad \mathbf{r}_{s(i,h)}(v_i) = \mathbf{r}_{s(i,j)}(v_i), \quad \forall h, j \in \text{ind}(\Gamma(v_i)), v_i \in V(\mathbf{G}),$$

$$(4.3) \quad \sum_h d_{is(i,h)} \{ p_{s(i,h)}^2 u_{s(i,h)}' \mathbf{e}_{s(i,h)} + q_{s(i,h)}^2 (w_{s(i,h)}' + \psi_{s(i,h)}) \mathbf{e}_{s(i,h)}^\perp + m_{s(i,h)}^2 \psi_{s(i,h)}' \mathbf{n} \} (\pi_{s(i,h)}^{-1}(v_i)) = \mathbf{g}_i, \quad \forall v_i \in V(\mathbf{G}).$$

$$(4.4) \quad \begin{aligned} \mathbf{r}_{s(i,h)}(v_i) &= 0, \quad \forall v_i \in \partial_0 V(\mathbf{G}), \\ \{ p_{s(i,h)}^2 u_{s(i,h)}' \mathbf{e}_{s(i,h)} + q_{s(i,h)}^2 (w_{s(i,h)}' + \psi_{s(i,h)}) \mathbf{e}_{s(i,h)}^\perp \} (\pi_{s(i,h)}^{-1}(v_i)) &= \mathbf{f}_i, \\ m_{s(i,h)}^2 \psi_{s(i,h)}' (v_i) &= h_i, \quad \forall v_i \in \partial_1 V(\mathbf{G}). \end{aligned}$$

In analogy to the previous section we define the operators

$$\begin{aligned} \mathbf{K}_j &:= p_j^2 \mathbf{P}_j + q_j^2 \mathbf{P}_j^\perp + m_j^2 \mathbf{P}_n, \\ \mathbf{M}_j &:= q_j^2 \langle \mathbf{P}_n(\cdot), \mathbf{n} \rangle \mathbf{e}_j^\perp, \\ \mathbf{N}_j &:= q_j^2 \langle \mathbf{P}_j^\perp(\cdot), \mathbf{e}_j^\perp \rangle, \mathbf{n} \\ \mathbf{S}_j &:= \mathbf{N}_j - \mathbf{M}_j \Rightarrow \mathbf{S}_j^* = -\mathbf{S}_j. \end{aligned}$$

Then the state equations (4.1) can be written as

$$(4.5) \quad \ddot{\mathbf{r}}_j - \mathbf{K}_j \mathbf{r}_j'' + \mathbf{S}_j \mathbf{r}_j' + q_j^2 \mathbf{P}_u \mathbf{r}_j = 0.$$

The corresponding continuity condition is exactly of the form as in (3.1), with the only difference that the vectors have now three components. Let us now consider the homogeneous situation. The balance of forces requirement can now be expressed as

$$(4.6) \quad \sum_h d_{ts(i,h)} (\mathbf{K}_{s(i,h)} \mathbf{r}'_{s(i,h)} + \mathbf{M}_{s(i,h)} \mathbf{r}_{s(i,h)}) (\pi_{s(i,h)}^\top)(v_i) = 0, \\ \forall v_i \in V(\mathbf{G}) \setminus \partial_0 V(\mathbf{G}).$$

## 5. Numerical Simulation of controlled networks

**5.1. Absorbing controls.** In order to give an exemplaric introduction, we will be dealing with the numerical realization of stabilizing feedback controls. Those controls are of more practical interest than open loop controls. In particular, the feedback controls we discuss are of the absorbing type. They can be viewed as being at the interface between controllability concepts and stabilizability concepts. Loosely speaking, an absorbing control is one which passes on the entire energy flux into one direction. The history of these controls goes back right to the beginning of control theory of distributed parameter systems and is connected with D.L. Russell, see [18] for an excellent review. From the mathematical viewpoint it is most easily explained for the 1-d wave equation in its characteristic description ([18]):

**Example 5.1.** *The 1-d-wave equation.*

Let us write down the scaled classical 1-d-wave equation as follows

$$\begin{cases} \ddot{w} = w'', & \text{on } (0, 1) \times (0, T), \\ w(0, t) = 0, \quad w'(1, t) = \phi, & t \in (0, T), \\ w(x, 0) = w_0(x), \quad \dot{w}(x, 0) = w_1(x), & x \in [0, 1]. \end{cases}$$

Under the transformations

$$\begin{aligned} u_1 &:= \dot{w}, \quad u_2 := w', \\ v_1 &:= \frac{1}{\sqrt{2}}(-u_1 + u_2), \quad v_2 := \frac{1}{\sqrt{2}}(u_1 + u_2), \end{aligned}$$

we have with  $\Gamma = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$  the first order system

$$\begin{aligned} \dot{\mathbf{v}} &= \Gamma \mathbf{v}', \\ v_1(0) &= v_2(0), \quad v_1(1) + v_2(1) = \sqrt{2}\phi. \end{aligned}$$

In this simple case the total energy and its time derivative are given by

$$E = \int_0^1 \|\mathbf{v}\|^2 dx, \quad \dot{E} = \frac{1}{2} \{-v_1(1)^2 + v_2(1)^2\}.$$

It is obvious that the optimal choice of a control is the one leading to  $v_2(1) = 0$  for all  $t \geq 0$ . In fact, it is readily seen that this choice amounts to cancelling the outgoing signals at the boundary point  $x = 1$ . This means that the control mimics the response of the half-infinite string. This is Russell's philosophy in [18]: "let nature take its course". In terms of feedback controls the goal of cancelling outgoing signals can be achieved by letting  $\phi(t) = \frac{1}{\sqrt{2}}v_1(1, t)$  which is equivalent to the closed loop boundary condition  $w'(1, t) = -\dot{w}(1, t)$ , which, in turn, is the classical *dead beat control*. Similar arguments can be applied to the Timoshenko system. We note, however, that because of the dispersion of waves, we do not have exact cancellation of outgoing signals at the boundary, to the extent that exact controllability by velocity feedback controls cannot be achieved. Nevertheless, exact controllability can clearly be proved for open-loop controls. It is apparent that besides the necessity of discretizing the state equation and the boundary conditions in appropriate way, the total energy as function of time plays a key role. In particular, *it appears to be necessary to use a numerical scheme which conserves the total energy of the discretized system as long as no controls are applied*. This can, for instance, be achieved by the classical *Lax-Friedrichs-scheme* discussed by Le Roux [17]. See also Halpern [9] for various schemes simulating absorbing boundaries. This scheme is applied to the system above in terms of the variables  $u_i$ , and it is also appropriate to quasi-linear models. To avoid unnecessary indexing, we write down  $u = w'$ ,  $v = \dot{w}$  where  $w$  solves

$$\ddot{w} = (\sigma(w'))'$$

for a suitable function  $\sigma$ , such that

$$\dot{u} = v', \quad \dot{v} = u'.$$

Upon introducing the grid  $x_i = i \cdot \Delta x$ ,  $i = 0, \dots, n$ ,  $t_j = j \cdot \Delta t$ ,  $j = 0, \dots, m$ , where  $\Delta x = 1/n$ ,  $\Delta t = q\Delta x$ ,  $q$  being the appropriate mesh ratio, we have ([17]):

$$(5.1) \quad \begin{aligned} w_{i+1}^{j+1} &= \frac{1}{2} \{(w_{i+1}^j + w_{i-1}^j) + \frac{q}{2}(v_{i+1}^j - v_{i-1}^j)\}, \\ v_{i+1}^{j+1} &= \frac{1}{2} \{(v_{i+1}^j + v_{i-1}^j) + \frac{q}{2}(\sigma(w_{i+1}^j) - \sigma(w_{i-1}^j))\}, \end{aligned}$$

and the boundary conditions are represented by

$$(5.2) \quad \begin{aligned} w_0^{j+1} &= 0, \quad w_n^{j+1} = 0, \\ w_0^{j+1} &= w_1^j + q \cdot v_1^j, \quad v_n^{j+1} = v_{n-1}^{j+1} - q \cdot w_{n-1}^j. \end{aligned}$$

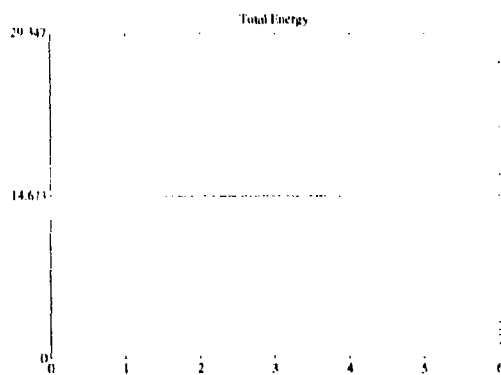


FIGURE 1. Conservation of energy

The Courant-Friedrichs-Lewy condition  $q \leq 1$ ; we always use  $q = 1$ . This set up is particularly useful if the boundary feedback above is to be implemented, because it prescribes the strain at the boundary  $x = 1$ . We simply put:  $w_n^{j+1} = -v_n^{j+1}$ . For further information, in particular on other realizations of absorbing boundary conditions see Halpern [9].

We first look at the uncontrolled string. We take 45 spatial mesh points and 225 time steps, which amounts to time span of  $[0, 5]$ . As initial conditions we choose  $w(x, 0) = \sin(3\pi(x)^2)$ ,  $\dot{w}(x, 0) = 0$ . The reason for this choice is that we do not want to look at eigenmodes but rather to a response to rich-enough data. Because of space restrictions, we display only the total energy, which is computed numerically using the trapezoidal rule, see figure 1. We are now going to apply the absorbing boundary control at the boundary  $x = 1$ . The result is shown in figure 2, where we display the strain, rather than the displacement, simply because it is the canonical variable in the Lax-Friedrichs-scheme. The total energy is displayed in figure 3.

These plots serve as evidence for the remarks above. The numerical procedure can be (and has been) applied also to planar strings and planar Timoshenko beams. For similar simulations, using however the classical direct discretization of the wave equation, see J. Schmidt [20].  $\square$

**5.2. Directing controls.** We pursue the concept of absorbing controls a bit further and consider now three identical strings and out-of-the-plane-displacement. The fact that we concentrate on this model, rather than on the planar system, is only for the sake of a convenient display of the plots. We will consider planar systems later.

**Example 5.2.** *Three serial strings.*

We have the following system:

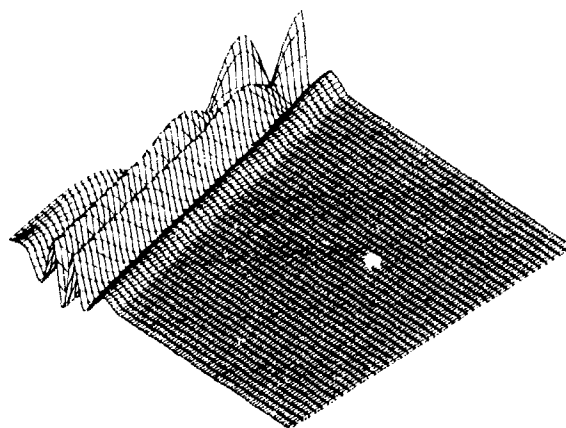


FIGURE 2. Absorbing boundary control

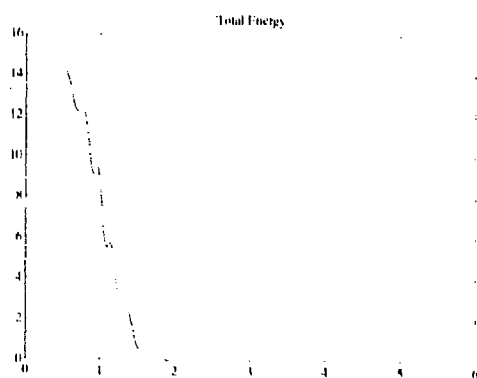


FIGURE 3. Energy plot of disappearing solution



$$\begin{aligned}
(5.3) \quad & \ddot{w}^i = (w^i)'' \quad \text{on} \quad (0, 1) \times (0, T), \\
& w^1(0) = 0, \quad w^1(1) = w^2(1), \quad (w^1)'(1) + (w^2)'(1) = f^1, \\
& w^3(1) = 0, \quad w^2(0) = w^3(0), \quad (w^2)'(0) + (w^3)'(0) = f^2, \\
& + \text{ initial conditions.}
\end{aligned}$$

This system describes the motion of three strings with obvious orientation of the running variable  $x$ . The orientation is chosen in order to avoid mixed boundary conditions. We perform a succession of transformations as above. Namely,

$$\begin{aligned}
u_1^i &:= \dot{w}^i, \quad u_2^i := (w^i)', \\
v_1^i &:= \frac{1}{\sqrt{2}}(-u_1^i + u_2^i), \quad v_2^i := \frac{1}{\sqrt{2}}(u_1^i + u_2^i), \\
z_i &:= v_1^i, \quad i = 1, 2, 3, \quad z_i := (v_2^{i-3})', \quad i = 4, 5, 6.
\end{aligned}$$

With these transformations we are able to write the system as

$$\begin{aligned}
\dot{z}^i &= -(z_i)', \quad i = 1, 2, 3, \\
\dot{z}^i &= (z_i)', \quad i = 4, 5, 6, \\
z_1(0) &= z_4(0), \quad z_1(1) - z_4(1) = z_2(1) - z_5(1), \\
z_3(1) &= z_6(1), \quad z_2(0) - z_5(0) = z_3(0) - z_6(0), \\
z_4(1) + z_2(1) &= \frac{1}{\sqrt{2}}f^1, \quad z_5(0) + z_3(0) = \frac{1}{\sqrt{2}}f^2,
\end{aligned}$$

together with initial conditions. Our goal now is to steer the energy of the entire system to the second string by applying absorbing controls at the two multiple joints connecting the second string with the first and the third string. We adopt the strategy of P.Hagedorn and J.Schmidt, see [20]. To this end, we define

$$(5.4) \quad f^1 := \sqrt{2}z_2(1), \quad f^2 := \sqrt{2}z_5(0).$$

By this choice the outgoing signals  $z_3(0)$  at  $x = 0$  and  $z_4(1)$  at the boundary  $x = 1$  are cancelled. For the sake of convenience, we write down the boundary conditions in matrix form also reflecting on incoming and outgoing signals as in Russell [18]:

$$\begin{aligned}
(5.5) \quad & \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} (0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_4 \\ z_5 \\ z_6 \end{pmatrix} (0), \\
& \begin{pmatrix} z_4 \\ z_5 \\ z_6 \end{pmatrix} (1) = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} (1).
\end{aligned}$$

We consider a particular exemplaric path of characteristics, namely, for  $t = 0$  we start with the three outgoing signals  $z_1, z_2, z_3$  moving to the boundary  $x = 1$  with characteristic speed equal to 1. These characteristics are exactly the incoming signals at that boundary. According to the reflection condition (5.5) at  $x = 1$ ,  $z_1$  is cancelled at  $x = 1$ . Now, the triple  $z_4, z_5, z_6$  constitutes the set of outgoing signals at  $x = 1$ , and accordingly, the set of incoming signals at  $x = 0$ . As  $z_4$  is already equal to zero at  $x = 0$  (because of the properties of characteristics), the set of nonvanishing outgoing signals at  $x = 0$  reduces to the variable  $z_2$ , according to the reflection condition (5.5). Therefore, after reflection and take off to the boundary at  $x = 1$ , the only nonvanishing incoming signal there is  $z_2$ , which, in turn, is reflected to  $z_5$  as the only nonvanishing outgoing signal at  $x = 1$ . This, however, closes the cycle, because  $z_5$  is converted to  $z_2$  upon reflection at  $x = 0$ , and so on. We depict this situation in the following diagram:

$$\begin{array}{ccc} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} & \xrightarrow{atx=1} & \begin{pmatrix} 0 \\ z_5 \\ z_6 \end{pmatrix} & \xrightarrow{atx=0} & \begin{pmatrix} 0 \\ z_2 \\ 0 \end{pmatrix} \\ & \xrightarrow{atx=1} & \begin{pmatrix} 0 \\ z_5 \\ 0 \end{pmatrix} & \xrightarrow{atx=0} & \begin{pmatrix} 0 \\ z_2 \\ 0 \end{pmatrix} \end{array}$$

This shows that after two reflections at the boundaries there is only a residual motion left in the second string. This is what is meant by directing or diode-type controls. We will demonstrate that concept for *planar* frames shortly. Let us first, however, provide some numerical evidence for the situation described above. In figure 4 we take an initial condition distributed along the three strings, which, incidentally, are chosen to have individual length  $\frac{1}{3}$ , so that the length of the entire system is equal to 1. That the energy remains constant after two reflections (that is at  $t = \frac{2}{3}$ ) is seen in figure 5.

In the context of the original system (5.3) the controls are chosen to be as follows:

$$f^1 = -\ddot{w}^2(1) + (w^2)'(1), \quad f^2 = \ddot{w}^2(0) + (w^2)'(0)$$

which results in the feedback controls

$$(w^1)'(1) = -\dot{w}^1(1), \quad (w^3)'(0) = \dot{w}^3(0).$$

It has been shown in [7] that the absorbing boundary feedback, the one with the exact impedance of the half-infinite string, is sensitive to time delays in the velocity feedback. In fact it has subsequently been shown by R. Datko [5], [6] that this phenomenon is not peculiar to strings, and that it is not even circumvented upon the introduction of severe damping. Let us show this instability for our three string model (5.3) in figure 6, and figure 7.

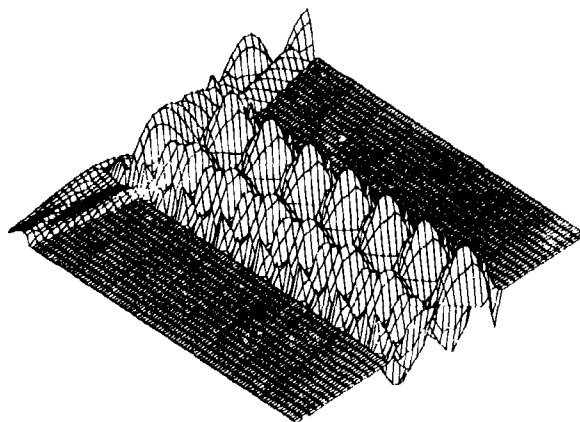


FIGURE 4. Directing controls

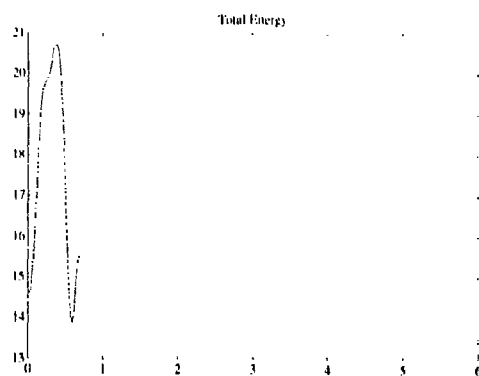


FIGURE 5. Energy of entire system

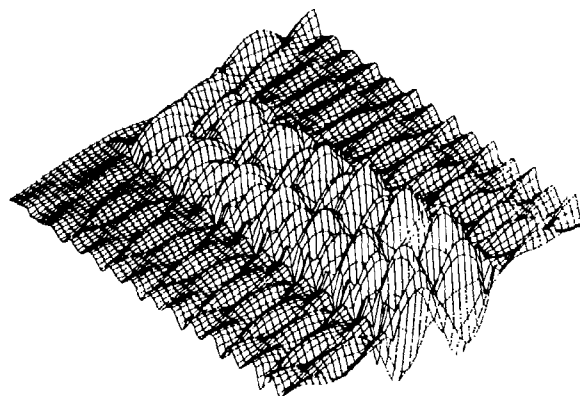


FIGURE 6. Effect of time-delay

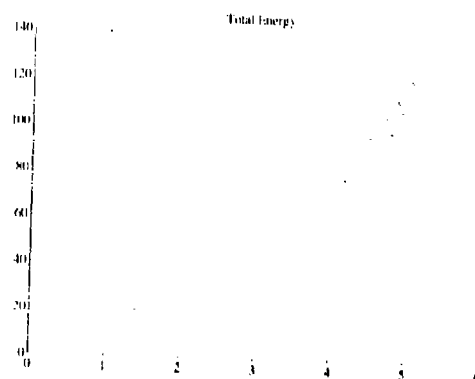


FIGURE 7. Instability due to time delay

Even though this instability is rather discouraging, the situation in a real structure is not as bad as this seems to suggest. First of all, some of the energy absorbing feedback controls can be implemented as passive devices, for which a time lag will not be present. But even if the feedback law is to be realized by numerical computation (in real time), the time-lag due to the action of the microprocessor will not be a constant, but, rather, distributed by a certain stochastic process. There is, however, another remedy to this problem: the growing instability of the solutions under a feedback with time-lag is caused by feeding back an increasing amount of control energy to the system, once it is out-of-phase. In a real structure, however, a controller will have to respect certain a priori given bounds. There is an extensive literature devoted to the problem of controllability/stabilizability by bounded controls. We do not want dwell on this question here. We refer the reader to a recent paper by Rao [16] and to the classical work of Haraux [10], [11] and also to Cabannes [2], [3]. Without going into the theoretical detail, we present the response of the system (5.3) without time delay but with respect to the nonlinear boundary controls:

$$f^1 = -\text{sign}(\dot{w}^2(1)) + (w^2)'(1), \quad f^2 = \text{sign}(\dot{w}^2(0)) + (w^2)'(0)$$

See figure 8, figure8. It is apparent that these controls lead to considerably improved decay to a residual motion in the second string. It is interesting to observe the pattern of the strain in that string. If we now take these nonlinear boundary controls and if we account for a small time delay in the control (say one time step) then we observe no such instability. We dispense with the mesh plot and display the energy plot only; see figure 8.

The proof that this observed robustness is actually true for the infinite dimensional system appears to be an open problem.  $\square$

The kind of nonlinearity of boundary controls just considered is in fact derived from a physical principle, namely *dry friction*. We come to this point later, when we are going to give an example of implicit Runge- Kutta methods applied to differential algebraic equations.

In the next example we want to demonstrate how directing controls can be used to improve the system response by 'directing' the energy to a 'safe' part of the structure. This is of great practical importance because a large flexible structure usually contains many uncontrollable circuits. Even if one puts the classical dissipative controls at each node, there will be a residual vibration in the system which will not decay to zero. Usage of 'directing' controls, however, can serve to 'channel' the energy through the network to end up at a set of nodes where either the energy influx is not important or where incoming energy is more easily absorbed.

**Example 5.3.** *Fine strings containing a circuit.*

This model consists of a unit square (labels 1,2,3,4) with another string (#5) attached to the node connecting string #2 and string #3. We have an initial

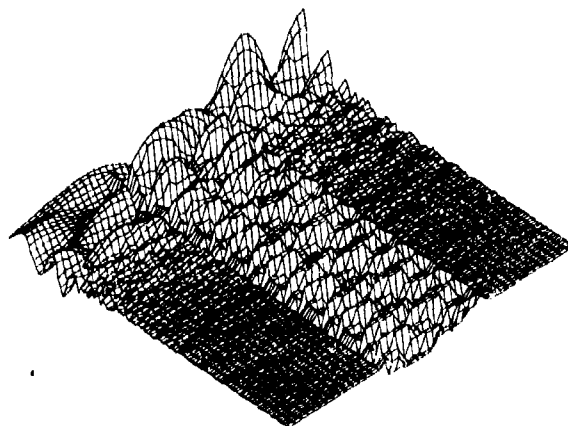
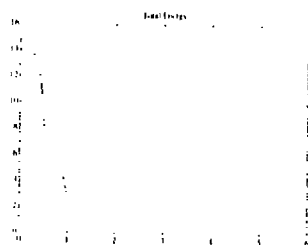
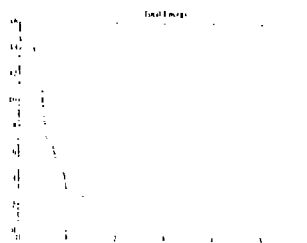


FIGURE 8. bang-bang control

FIGURE 9  
Improved decayFIGURE 10  
Robustness

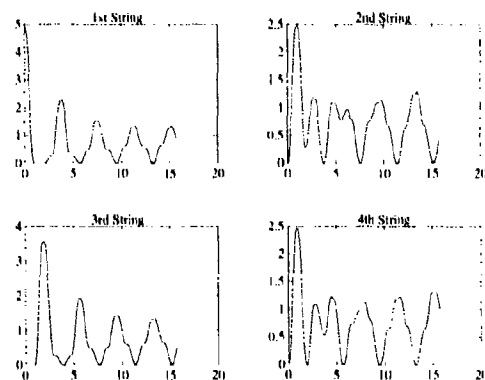


FIGURE 11. Energy in individual strings

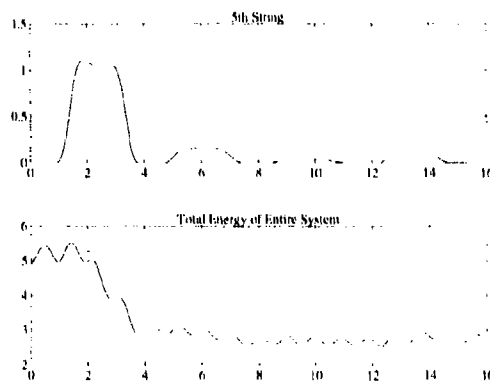


FIGURE 12. Energy in 5th. string and entire system

displacement in the first string only (as usual). In the first experiment we apply an absorbing control only at the free end of the attached string.

Figures 11,12 reveal what one expects.

Energy is taken out of the circuit until a residual motion settles. In particular, the plot of the total energy of the 5th string shows that with each cycle less energy can be absorbed from the circuit because of the gradually vanishing of the nodal excitation. The situation completely changes, once 'directing' controls are introduced around the vertex connecting strings #2, #3 and #5. See figures 13 and 14.

□

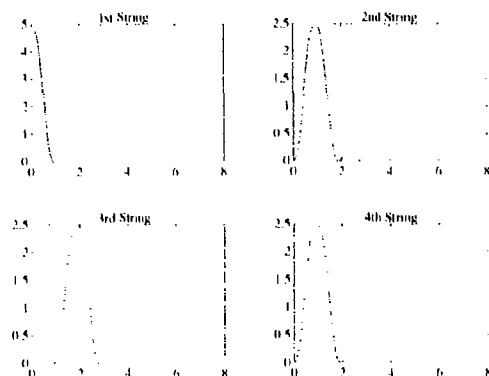


FIGURE 13. Directing controls

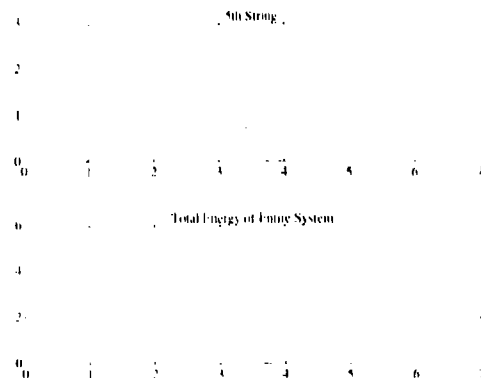


FIGURE 14

**5.3. Finite elements.** As mentioned above, the previous calculations have been based on string models. In dealing with networks of Timoshenko beams the numerical simulation becomes more involved, and it appears less appealing to use finite difference schemes. In the case of networks of Timoshenko beams and also for more complex string networks, the finite-element method seems best suited. Because of space limitations, we have to be very brief and refer the reader to Hughes [12]. Our concern in this section is to give numerical evidence for the lack of approximate controllability / stabilizability of a network of planar Timoshenko beams containing a circuit.

Let  $\mathbf{K}$ ,  $\mathbf{C}$ ,  $\mathbf{M}$  denote the global stiffness, global damping and global mass matrix, respectively. Then the system of equations governing the evolution of the



time dependent nodal variables can be written down as follows.

$$(5.6) \quad \mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\dot{\mathbf{U}} + \mathbf{K}\mathbf{U} = \mathbf{F},$$

where we use  $\mathbf{U}$  as the vector containing the nodal variables of all element nodes ( $\mathbf{F}$  represents all applied forces). We recall the family of Newmark schemes which are actually of the Predictor-Corrector type. Again, we follow Hughes [12]. To this end we put

$$(5.7) \quad \mathbf{a}_n = \ddot{\mathbf{U}}(t_n), \quad \mathbf{v}_n = \dot{\mathbf{U}}(t_n), \quad \mathbf{d}_n = \mathbf{U}(t_n),$$

for the acceleration, velocity and displacement, respectively. We then define the *predictors* by

$$(5.8) \quad \begin{aligned} \tilde{\mathbf{d}}_{n+1} &= \mathbf{d}_n + \Delta t \mathbf{v}_n + \Delta t^2 \frac{(1-2\beta)}{2} \mathbf{a}_n, \\ \tilde{\mathbf{v}}_{n+1} &= \mathbf{v}_n + (1-\gamma) \Delta t \mathbf{a}_n. \end{aligned}$$

As for starting values, we have

$$\mathbf{M}\mathbf{a}_0 = \mathbf{F} - \mathbf{C}\mathbf{v}_0 - \mathbf{K}\mathbf{d}_0.$$

This equation can be solved for  $\mathbf{a}_0$ . In the step  $n+1$  the acceleration  $\mathbf{a}_{n+1}$  is calculated from previous quantities and a priori known data by the recursion

$$(5.9) \quad (\mathbf{M} + \gamma \Delta t \mathbf{C} + \beta \Delta t^2 \mathbf{K}) \mathbf{a}_{n+1} = \mathbf{F}_{n+1} - \mathbf{C} \tilde{\mathbf{v}}_{n+1} - \mathbf{K} \tilde{\mathbf{d}}_{n+1}.$$

Once  $\mathbf{a}_{n+1}$  is known from (5.9), we update the predictors via so-called *correctors*

$$(5.10) \quad \begin{aligned} \mathbf{d}_{n+1} &= \tilde{\mathbf{d}}_{n+1} + \beta \Delta t^2 \mathbf{a}_{n+1}, \\ \mathbf{v}_{n+1} &= \tilde{\mathbf{v}}_{n+1} + \gamma \Delta t \mathbf{a}_{n+1}. \end{aligned}$$

The result of (5.10) is then inserted into (5.8) which, in turn, will then be used in (5.9) and so on. The two Newmark parameters can be chosen in various ways resulting in various different properties of the scheme. We will always make the choice:  $\beta = \frac{1}{4}$ ,  $\gamma = \frac{1}{2}$ . This choice is known to lead to conservation of the energy above.

In our calculations we confine ourselves to planar frames of Timoshenko beams.

In that case the local stiffness matrix  $\mathbf{k}^e$  is a  $6 \times 6$  matrix as we have vertical and longitudinal displacement and one rotation in the plane for each of the two nodes of one element. For the sake of completeness we write it out explicitly

$$(5.11) \quad \mathbf{k}^c = \begin{pmatrix} \frac{EA}{\ell} & 0 & 0 & -\frac{EA}{\ell} & 0 & 0 \\ 0 & \frac{A^*}{\ell} & \frac{A^*}{2} & 0 & -\frac{A^*}{\ell} & \frac{A^*}{2} \\ 0 & \frac{A^*}{\ell} & \frac{A^*\ell}{4} + \frac{EI}{2} & 0 & -\frac{A^*}{\ell} & \frac{A^*\ell}{4} - \frac{EI}{2} \\ -\frac{EA}{2} & 0 & 0 & \frac{EA}{\ell} & 0 & 0 \\ 0 & -\frac{A^*}{\ell} & -\frac{A^*}{2} & 0 & \frac{A^*}{\ell} & -\frac{A^*}{2} \\ 0 & \frac{A^*}{\ell} & \frac{A^*\ell}{4} - \frac{EI}{2} & 0 & -\frac{A^*}{\ell} & \frac{A^*\ell}{4} + \frac{EI}{2} \end{pmatrix}$$

As for the lumped mass matrix we take

$$(5.12) \quad \mathbf{m}^c = \frac{m_0 A \ell}{2} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{I}{A} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{I}{A} \end{pmatrix}$$

See [12] p. 515 for the planar beam without longitudinal displacement. With the bar wave velocity  $c = \sqrt{E/m_0}$ , and the beam shear wave velocity  $c_s = \frac{GA^*}{m_0 A}$ , the critical time step  $\Delta t$  can be estimated as

$$\Delta t \leq \min\left\{\frac{h}{c}, \left(\frac{h}{c_s}\right)\left(1 + \frac{A}{I}\left(\frac{h}{2}\right)^2\right)^{-\frac{1}{2}}\right\}.$$

**Example 5.4.** *The unit square of planar Timoshenko beams*

Here we want to show two things: at first we consider the uncontrolled system. Again, we have an initial displacement only in the bottom beam (#1), which is horizontal. The difficulty in showing all the displacements and rotations is obvious. At the nodes local vertical motion of beam #1 converts to local longitudinal motion of the upright beams #2 and #4, which, in turn, is converted to local vertical motion in the top (horizontal) beam(#3). Because in the finite-element approach we automatically have the local rotations  $\Gamma_i$  available, we can easily go back to the global reference frame. In this way the locally longitudinal motion of the 2nd and 4th beam is, indeed, globally vertical. See figure 15. This plot clearly shows the collision of waves in the 3rd beam after travelling safely through beams #2 and #4. The initial bump next to the corner at the bottom indicates the location of the first beam. Figure 16 shows the conservation of energy.

In the second experiment we apply dissipative controls at all four nodes. See figure 17 and figure 18. Figure 17 clearly shows how the motion settles at a residual oscillation with nodes at the vertices. Accordingly, the energy settles at a constant value; see figure 18.

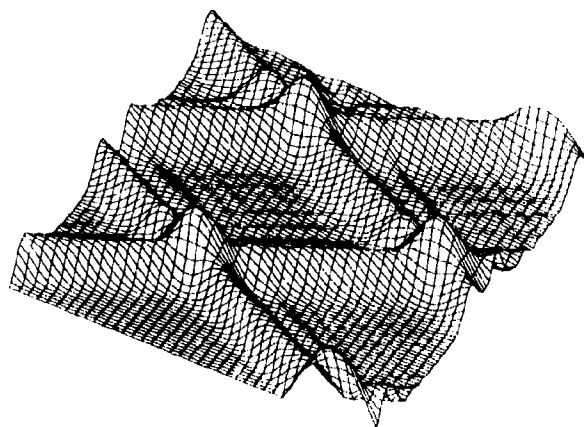


FIGURE 15. Unit square: uncontrolled motion

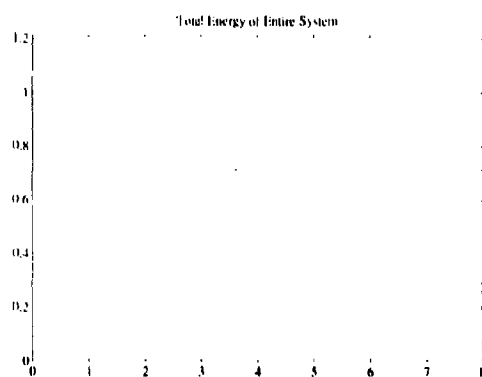


FIGURE 16. Unit square: conservation of energy

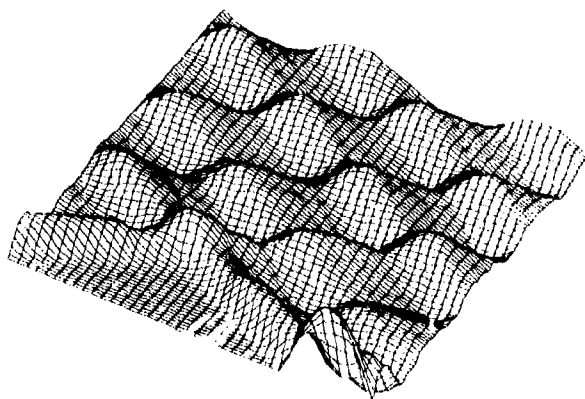


FIGURE 17. Unit square: controlled motion

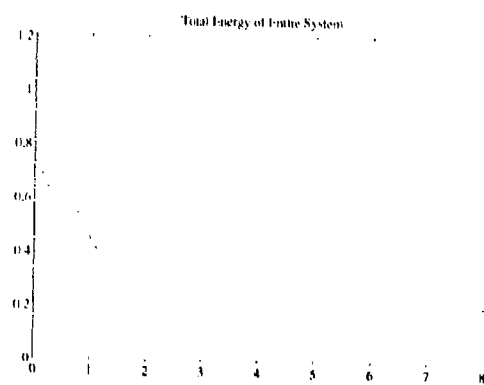


FIGURE 18. Unit square: saturation of energy

## 6. Implicit Runge-Kutta method applied to differential algebraic systems in the context of dry friction at the joints

We consider two identical strings of length 1 put together at the end point where  $x = 1$  for both strings. We assume clamped boundaries at  $x = 0$  for both strings, and we assume dry friction at the common point  $x = 1$ . The resulting model is as follows:

$$(6.1) \quad \begin{aligned} \ddot{w}^i &= (w^i)'', \quad i = 1, 2, \\ w^1(0) &= w^2(0) = 0, \quad w^1(1) = w^2(1), \\ (w^1)'(1) + (w^2)'(1) &= -\rho_s z, \quad z = \begin{cases} \text{sign}(\dot{w}^1(1)), & \text{if } \dot{w}^1(1) \neq 0 \\ \in [-\frac{\rho_a}{\rho_s}, \frac{\rho_a}{\rho_s}], & \text{else} \end{cases}, \\ w^1(x, 0) &= w_0^1(x), \quad w^2(x, 0) = 0, \quad \dot{w}^1(x, 0) = \dot{w}^2(x, 0) = 0. \end{aligned}$$

Existence and uniqueness of solutions to this and much more general systems can be treated along the lines of Haraux [10]. In this special case we use the methods of characteristics. We perform the familiar transformations

$$\begin{aligned} u_1 &:= \dot{w}^1, \quad u_2 := \dot{w}^2, \quad u_3 := (w^1)', \quad u_4 := (w^2)', \\ v_1 &:= \frac{1}{\sqrt{2}}(-u_1 + u_3), \quad v_3 := \frac{1}{\sqrt{2}}(u_1 + u_3), \\ v_2 &:= \frac{1}{\sqrt{2}}(-u_2 + u_4), \quad v_4 := \frac{1}{\sqrt{2}}(u_2 + u_4). \end{aligned}$$

We then have the system

$$\dot{\mathbf{v}} = \mathbf{\Gamma} \mathbf{v}', \quad \mathbf{\Gamma} = \text{diag}([-1, -1, 1, 1]),$$

with the boundary and node conditions

$$\begin{aligned} (v_3 - v_1)(0) &= 0, \quad (v_3 - v_1)(1) = (v_4 - v_2)(1), \\ (v_4 - v_2)(0) &= 0, \quad (v_3 + v_1)(1) + (v_4 + v_2)(1) = -\rho_s \sqrt{2} z, \end{aligned}$$

with

$$z = \begin{cases} \text{sign}(v_3 - v_1)(1) & \text{if } (v_3 - v_1)(1) \neq 0 \\ \in [-\frac{\rho_a}{\rho_s}, \frac{\rho_a}{\rho_s}], & \text{else.} \end{cases}$$

These can be rewritten as

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} (0) = \begin{pmatrix} v_3 \\ v_4 \end{pmatrix} (0), \quad \begin{pmatrix} v_3 \\ v_4 \end{pmatrix} (1) = - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} (1) - \rho_s \frac{1}{\sqrt{2}} z \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

As a consequence of our special choice of initial conditions (we could do more general but we want to treat this problem only as an exemplaric one), we have  $v_2(\cdot, 0) = v_4(\cdot, 0) = 0$ .

We will consider two cases:

- i.)  $|(w^1)'(\cdot, 0)| \leq \rho_a$
- ii.)  $|(w^1)'(\cdot, 0)| > \rho_a$

*We claim that in case i.) no motion takes place on the second string for  $t \geq 0$ .*

*Proof.* We have

$$(v_3 - v_1)(1) = 0 \text{ at } t = 0.$$

Take

$$z = -\frac{\sqrt{2}}{\rho_s} v_1(1, t) \in \left[-\frac{\rho_a}{\rho_s}, \frac{\rho_a}{\rho_s}\right], \quad (t \in [0, 1]).$$

We follow the reflections at  $x = 1$  and subsequently at  $x = 0$ , i.e.

$$\begin{aligned} v_3(1, t) &= -v_{20}(1 - t) + v_{10}(1 - t) = v_{10}(1 - t), \\ v_4(1, t) &= 0, \quad t \in [0, 1]. \end{aligned}$$

$$\begin{aligned} v_1(0, 1 + t) &= v_{10}(1 - t), \\ v_2(0, 1 + t) &= 0. \end{aligned}$$

Hence, upon reflection,  $v_1$  does not change magnitude, and with  $z$  as above,  $v_4$  is kept equal to zero, causing also  $v_2$  to be equal to zero.

The analogous case starting with  $v_{30}, v_{40}$  and travelling towards the boundary at  $x = 0$  first is settled in a similar way. The main point is that the variable  $z$  (representing the adhesive force) always compensates the shear force ('kills' the outgoing signal  $v_4(1, t)$ ) at the boundary  $x = 1$  which represent the multiple node in our original model.  $\square$

*We also claim that in case ii.) motion does spill over into the second string*

*Proof.* We have initially  $(v_3 - v_1)(1, 0) = 0$ . Assume now

$$(v_3 - v_1)(1, t) = 0, \quad t \in (0, \varepsilon).$$

$$\begin{aligned} z(t) \in \left[-\frac{\rho_a}{\rho_s}, \frac{\rho_a}{\rho_s}\right] &\Rightarrow z(t) \neq -\frac{\sqrt{2}}{\rho_s} v_1(1, t), \text{ on } (0, \varepsilon) \\ &\Rightarrow v_4(1, t) \neq 0, \text{ on } (0, \varepsilon) \Rightarrow v_2(0, 1 + t) \neq 0. \end{aligned}$$

The latter inclusion says that there is non zero motion in the second string. The other cases are handled in a similar way.  $\square$

We are going to give numerical evidence to these remarks. We consider the mesh points  $x_i = \frac{i}{N}$ ,  $N = 40$  and use the classical finite-difference approximation to the second order operator (which, incidentally, equals the corresponding finite

element stiffness matrix). Indeed, we can view this process happening in one string with dry friction at the midpoint. We have the following system:

$$(6.2) \quad \ddot{w}_i - N^2 \{w_{i-1}(t) - 2w_i(t) + w_{i+1}(t)\} = \begin{cases} -\rho_s z & \text{if } i = \frac{N}{2} \\ 0 & \text{else} \end{cases}$$

where now  $z$  is given by

$$(6.3) \quad \begin{cases} \text{if } \dot{w}_{\frac{N}{2}}(t) \neq 0 & \Rightarrow z = \text{sign}(\dot{w}_{\frac{N}{2}}(t)) \\ \text{if } \dot{w}_{\frac{N}{2}}(t) = 0 & \Rightarrow \rho_s z \in \rho_n[-1, 1]. \end{cases}$$

We also have boundary and initial conditions:

$$(6.4) \quad w_0(t) = 0, w_N(t) = 0, w_i(0) = \dot{w}_i(0), \dot{w}_i = 0.$$

In this system, the variable  $z$  can be viewed as a control variable which forces the the solution of the system (6.2) to satisfy  $\dot{w}_{\frac{N}{2}} = 0$ . In order to do that, this variable  $z$ , which physically represents the adhesive force, will vary in time. In order to make the rôle of  $z$  more transparent, we consider a general differential algebraic equation of the following kind

$$\begin{cases} \dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}, \mathbf{z}), \\ 0 = \mathbf{g}(\mathbf{y}). \end{cases}$$

where  $\mathbf{f} : \mathbf{R}^n \times \mathbf{R}^m \longrightarrow \mathbf{R}^n$ ,  $\mathbf{g} : \mathbf{R}^n \longrightarrow \mathbf{R}^m$  are  $C^1, C^2$ -functions, respectively. Denote the Jacobian of  $\mathbf{f}$  with respect to  $\mathbf{y}, \mathbf{z}$  by  $\mathbf{f}_{\mathbf{y}}, \mathbf{f}_{\mathbf{z}}$ . Then we may differentiate (6.5)<sub>2</sub> with respect to the variable  $x$  we obtain

$$0 = \mathbf{g}_{\mathbf{y}}(\mathbf{y})\mathbf{f}(\mathbf{y}, \mathbf{z}).$$

Another differentiation with respect to  $x$  yields

$$(6.5) \quad 0 = (\mathbf{g}_{\mathbf{y}\mathbf{y}}(\mathbf{y})\mathbf{f}(\mathbf{y}, \mathbf{z}) + \mathbf{g}_{\mathbf{y}}(\mathbf{y})\mathbf{f}_{\mathbf{y}}(\mathbf{y}, \mathbf{z}))\mathbf{f}(\mathbf{y}, \mathbf{z}) + \mathbf{g}_{\mathbf{y}}(\mathbf{y})\mathbf{f}_{\mathbf{z}}(\mathbf{y}, \mathbf{z})\mathbf{z}'.$$

Obviously, if  $\mathbf{g}_{\mathbf{y}}(\mathbf{y})\mathbf{f}_{\mathbf{z}}(\mathbf{y}, \mathbf{z})$  is invertible in a neighbourhood of a solution, say  $(\bar{\mathbf{y}}, \bar{\mathbf{z}})$  then (6.5) is equivalent to an ordinary differential equation. This situation is said to have index 2. Also, in this case for the problem

$$\mathbf{G}(\mathbf{y}, \mathbf{z}) := \mathbf{g}_{\mathbf{y}}(\mathbf{y})\mathbf{f}(\mathbf{y}, \mathbf{z}) = 0, \quad \mathbf{G}(\bar{\mathbf{y}}, \bar{\mathbf{z}}) = 0,$$

by the implicit function, there exists a neighbourhood  $\mathcal{U}(\bar{\mathbf{y}}, \bar{\mathbf{z}})$  and a solution  $\mathbf{z}(\mathbf{y}, x)$  of that equation such that (6.5) holds. See Hairer and Wanner [8] for the theory of differential algebraic equations. With this remark in mind we go back to our system, where the case just considered corresponds to (6.2).(6.3)<sub>2</sub>.

If, now,  $z$  is in the mode 2 of (6.3) and reaches the value  $\frac{\rho_n}{\rho_s}$ , one has to switch to the first case in (6.3), which together with (6.2) then reduces to an ordinary differential equation. As a result, we have to switch between a differential

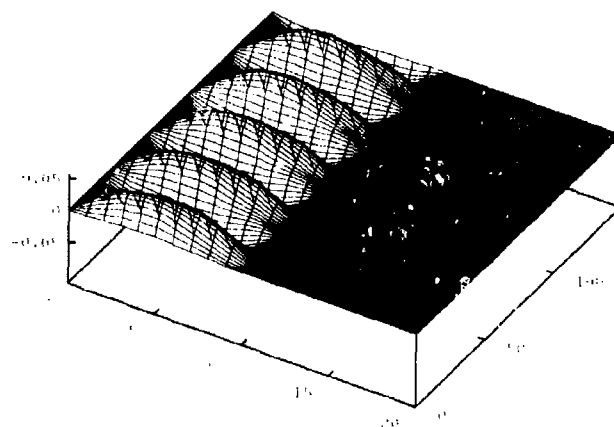


FIGURE 19. Dry friction: locking

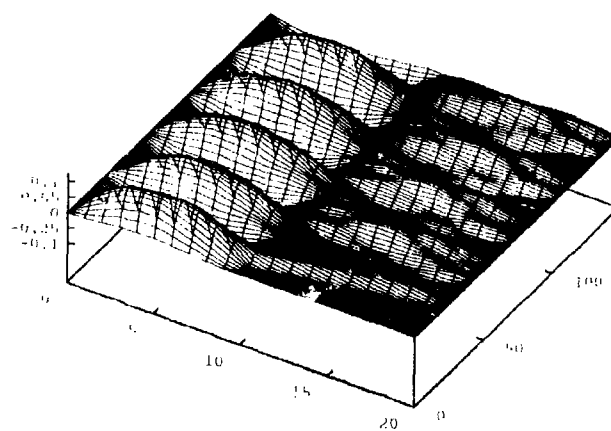


FIGURE 20. Dry friction: spill over



algebraic system (6.2), (6.3)<sub>2</sub> and (6.4) and an ordinary differential equation (initial value problem) (6.2), (6.3)<sub>1</sub> and (6.4). Switching problems of this kind have been treated by Bracey and Hairer [1]. They used the half- explicit 5 -stage Runge-Kutta method of order 4. We used the Radau5- implementation of the Radau IIa fully implicit Runge-Kutta method of order 5 also given by Hairer and Wanner in [8]. The corresponding driver was originally written by J. Verscht (Bayreuth) for a different problem. We made the appropriate changes to use it also for the problem under consideration and also for beam applications (which we do not reproduce here). The problem of switching between the two models is a major difficulty which is usually solved using 'dense output formulae' and some kind of extrapolation to find the switching manifold  $\mathbf{g}(\mathbf{y}) = 0$ . See the figures 20,19.

### References

1. V. Bracey and E. Hairer, *Half-explicit Runge-Kutta methods for differential-algebraic systems of index 2*, SIAM J. Numer. Anal. **30** (1993), no.2, 538-552.
2. H. Cabannes, *Mouvement d'une corde vibrante soumise à un frottement solide*, C. R. Acad. Sci. Paris **287** (1978), 671-673.
3. ———, *Propagation des discontinuités dans les cordes vibrantes soumises à un frottement solide*, C. R. Acad. Sci. Paris **289** (1979), 127-130.
4. G. Chen and Coleman, M. and H.H. West, *Pointwise stabilization in the middle of the span for second order systems, nonuniform and uniform exponential decay of solutions*, SIAM J. Control Optim. **47** (1987), no.4, 751-779.
5. R. Datko, *Not all feedback stabilized hyperbolic systems are robust with respect to small time delays in their feedbacks*, SIAM J. Control Optim. **26** (1988) 697-713.
6. ———, *A rank-one perturbation result on the spectra of certain operators* Proceedings of the Royal Society of Edinburgh **111A** (1989), 337-346.
7. R. Datko and Lagnese, J. E. and M. P. Polis, *An example of the effect of time delays in boundary feedback stabilization of wave equations* SIAM J. Control Optim. **24** (1986), 152-156.
8. E. Hairer and G. Wanner, *Solving ordinary differential equations II. Stiff and differential algebraic systems*, Springer-Verlag, Berlin, 1987.
9. L. Halpern, *Absorbing boundary conditions for the discretization of the one-dimensional wave equation*, Mathematics of Computation **38** (1982), no.158, 415-429.
10. A. Haraux, *Comportement à l'infini pour une équation d'ondes non linéaire dissipative*, C. R. Acad. Sci. Paris **287** (1978), 507-509.
11. ———, *Nonlinear evolution equations - global behavior of solutions*, Lect. Notes in Mathematics, vol.841, Springer-Verlag, Berlin, 1981.
12. T. J. R. Hughes, *The finite element method. Linear static and dynamic finite element analysis*, Prentice-Hall, New Jersey, 1987.
13. J. E. Lagnese and Leugering, G. and E. J. P. G. Schmidt, *Controllability of planar networks of Timoshenko beams*, SIAM J. Control. and Optim. **31** (1993), 780-811.
14. ———, *Modelling, analysis and control of multi-link flexible structures*, Birkhäuser Verlag, 1994.
15. G. Leugering and E. J. P. G. Schmidt, *On the control of networks of vibrating strings and beams*, Proceedings of the 28th IEEE Conference on Decision and Control, vol.3, IEEE, 1989, pp. 2287-2290.
16. Bopeng Rao, *Stabilisation du modèle SCOLE par un contrôle a priori borné*, C.R. Acad. Sci. Paris, **316** (1993), 1061-1066.

17. A. Y. Le Roux, *Stability of numerical schemes solving quasi-linear wave equations*, Mathematics of Computation **36** (1981), no.153, 93-105.
18. D. L. Russell, *Controllability and stabilizability theory for linear partial differential equations: recent progress and open questions*, SIAM Review **20** (1978), 639-739.
19. E. J. P. G. Schmidt, *On the modelling and exact controllability of networks of vibrating strings*, SIAM J. of Control and Optimization **30** (1992), 229-245.
20. J. Schmidt, *Entwurf von Reglern zur aktiven Schwingungsdämpfung an flexiblen mechanischen Strukturen*, Ph.D. thesis, 1987.

Günter Leugering  
Mathematisches Institut  
Universität Bayreuth  
D 95440 Bayreuth, Germany  
E-mail: leugering@uni-bayreuth.de

## VARIOUS RELAXATIONS IN OPTIMAL CONTROL OF DISTRIBUTED PARAMETER SYSTEMS

TOMÁŠ ROUBÍČEK

Institute of Information Theory and Automation  
Czech Academy of Sciences

**ABSTRACT.** The so-called convex-compactification theory is applied to an extension (=relaxation) of optimal control problems involving evolution distributed parameter systems. An infinite number of relaxed problems and corresponding Pontryagin maximum principles are thus obtained, including those described in the literature. A comparison and an abstract unifying viewpoint is thus made possible.

**1991 Mathematics Subject Classification.** 49J20, 49K20

**Key words and phrases.** Convex compactifications, optimal control, relaxation, Young measures, Pontryagin maximum principle, parabolic problems.

### 0. Introduction.

Optimal control problems, game-theoretical problems, or variational problems in their usual formulation typically do not admit any solution because of the absence of any compact topology making the data of the problem continuous. This is related in general with a lack of smoothness, or oscillation or concentration phenomena. Especially the oscillation phenomenon (i.e. the " $\varepsilon$ -optimal" controls necessarily oscillate more and more rapidly when  $\varepsilon \searrow 0$ ) appears typically in nonconvex and/or nonlinear optimal control problems. Therefore, an extension (=compactification) of the original space of admissible controls is urgently needed. This procedure, called "relaxation", has many variants especially in the cases where controls are functions of several variables (e.g. space and time), which appears typically in control of evolution distributed parameter systems. The aim of this paper is to present briefly a general theory, which covers by a unified way various relaxations that appeared in the literature, and to compare them together with the resulting Pontryagin maximum principle which is intimately related with the chosen relaxation.

# 1. Convex compactification theory.

Let us first consider an abstract topological space  $U$  of "original controls". The compactification of  $U$  (i.e. a continuous dense imbedding of  $U$  into a compact space  $K$ ) is an inappropriately general approach for our purposes, so that we impose a restrictive requirement that  $K$  is to be a convex subset of a locally convex space  $Z$ . Then we will speak about a convex compactification. More precisely, a triple  $(K, Z, i)$  will be called a convex compactification of  $U$  if

- (1a)  $K$  is a convex compact subset of  $Z$ ,
- (1b)  $Z$  is a locally convex space,
- (1c)  $i : U \rightarrow K$  is continuous,
- (1d)  $i(U)$  is dense in  $K$ .

Mostly, we will deal with Hausdorff convex compactifications (i.e.  $i$  injective) and then we will often identify  $U$  with  $i(U)$  for simplicity.

To compare various convex compactifications of  $U$ , we introduce a natural ordering: for  $(K_1, Z_1, i_1)$  and  $(K_2, Z_2, i_2)$  two convex compactifications of  $U$ , we will say that  $(K_1, Z_1, i_1)$  is finer than  $(K_2, Z_2, i_2)$  if there is a continuous affine surjection  $\psi : K_1 \rightarrow K_2$  such that  $\psi \circ i_1 = i_2$ . Then we will also say that  $(K_2, Z_2, i_2)$  is coarser than  $(K_1, Z_1, i_1)$ , and write  $(K_1, Z_1, i_1) \succeq (K_2, Z_2, i_2)$ , or briefly  $K_1 \succeq K_2$ . If  $(K_1, Z_1, i_1)$  is simultaneously finer and coarser than  $(K_2, Z_2, i_2)$ , then we will say that they are equivalent to each other, and write  $K_1 \cong K_2$ . If  $K_1 \succeq K_2$  but  $K_1 \not\cong K_2$ , then we will say that  $K_1$  is strictly finer than  $K_2$ , and write  $K_1 \succ K_2$ .

Having a convex compactification  $(K, Z, i)$ , we will sometimes omit  $Z$  and  $i$  when clear for the context. Examples of convex compactifications will be given later.

On "canonical" construction is always possible. Let us denote by  $C(U)$  a Banach space of continuous bounded functions on  $U$ , and let us consider a linear subspace  $\mathcal{F} \subset C(U)$  containing constants and satisfying:  $\forall u_1, u_2 \in U \exists$  a net  $\{u_\alpha\}$   $\forall f \in \mathcal{F} : \lim f(u_\alpha) = \frac{1}{2}f(u_1) + \frac{1}{2}f(u_2)$ ; such subspaces will be called "convexifying". Besides, let  $e : U \rightarrow \mathcal{F}^* : u \mapsto (f \mapsto f(u))$  denote the evaluation mapping, let the dual space  $\mathcal{F}^*$  be endowed with the weak\* topology, and let

$$M(\mathcal{F}) = \{\mu \in \mathcal{F}^* ; \|\mu\|_{\mathcal{F}^*} = 1 \text{ \& } \langle \mu, 1 \rangle = 1\}.$$

The elements of  $M(\mathcal{F})$  are usually called means on  $\mathcal{F}$ . The following assertion summarizes some results from [20, 24].

**Theorem 1.** *Let  $\mathcal{F}$  be a subspace of  $C(U)$  containing constants. Then  $(M(\mathcal{F}), \mathcal{F}^*, e)$  is a convex compactification of  $U$  if and only if  $\mathcal{F}$  is convexifying. Moreover, every convex compactification is equivalent with  $M(\mathcal{F})$  for some closed convexifying subspace  $\mathcal{F} \subset C(U)$ . If  $\mathcal{F}_1, \mathcal{F}_2$  are two convexifying subspaces of  $C(U)$  containing constants, then  $M(\mathcal{F}_1) \succeq M(\mathcal{F}_2)$  provided  $\mathcal{F}_1 \supset \mathcal{F}_2$ . If the closures of  $\mathcal{F}_1$  and  $\mathcal{F}_2$  in  $C(U)$  coincide with each other, then  $M(\mathcal{F}_1) \cong M(\mathcal{F}_2)$ . Conversely, if  $\mathcal{F}_1 \supset \mathcal{F}_2$  but their closures in  $C(U)$  do not coincide, then  $M(\mathcal{F}_1) \succ M(\mathcal{F}_2)$ .*

Finally, every convexifying subspace  $\mathcal{F} \subset C(U)$  is contained in some maximal convexifying subspace.

For a convex compactification  $K$  of  $U$ , the equivalent convex compactification  $M(\mathcal{F})$  will be referred to as a canonical form of  $K$ .

## 2. An abstract relaxation pattern.

If one neglects a concrete structure, every optimal control problem eventually takes the form of an abstract minimization problem:

$$(P) \quad \text{Minimize } \Phi(u) \text{ over } u \in U,$$

where  $\Phi : U \rightarrow \mathbb{R}$  is a cost function. As mentioned in Sect. 0, (P) need not have any solution and the need of its relaxation by a compactification of  $U$  and an extension of  $\Phi$  immediately arises. The reader can certainly anticipate that we will admit only a convex compactification of  $U$ , and the canonical form will be used. Therefore, the relaxed problem takes the form

$$(RP) \quad \text{Minimize } \tilde{\Phi}(\mu) \text{ over } \mu \in M(\mathcal{F}),$$

where  $(M(\mathcal{F}), \mathcal{F}^*, c)$  is some convex compactification of  $U$  and  $\tilde{\Phi} : M(\mathcal{F}) \rightarrow \mathbb{R}$  a l.s.c. (=lower semicontinuous) extension of  $\Phi$ ; i.e.  $\tilde{\Phi} \circ c = \Phi$ .

It is obvious that (RP) possesses always a solution, which is to be considered as a generalized solution of (P). Indeed, the relation between (RP) and (P) is very intimate: every cluster point in  $\mathcal{F}^*$  of  $\{c(u_k)\}_{k \in \mathbb{N}}$ , where  $\{u_k\}_{k \in \mathbb{N}}$  is a minimizing sequence for (P), solves (RP) and conversely every solution of (RP) can be reached by a net  $\{c(u_\alpha)\}$  such that  $\lim_\alpha \Phi(u_\alpha) = \inf \Phi(U)$ . In particular, every solution of (P) solves (after being imbedded via  $c$ ) the relaxed problem (RP).

The convex structure enables to set up the first-order necessary optimality condition. For this reason, we will say that  $\tilde{\Phi}$  has at  $\mu_0 \in M(\mathcal{F})$  the Gâteaux differential  $\tilde{\Phi}'(\mu_0) \in \mathcal{F}$  if

$$(2) \quad \forall \mu \in M(\mathcal{F}) : \quad \langle \mu - \mu_0, \tilde{\Phi}'(\mu_0) \rangle = \lim_{\varepsilon \searrow 0} \frac{1}{\varepsilon} (\tilde{\Phi}(\mu_0 + \varepsilon(\mu - \mu_0)) - \tilde{\Phi}(\mu_0)).$$

Let us only remind that, since  $\mathcal{F}^*$  is endowed with the weak\* topology, every linear continuous functional on  $\mathcal{F}^*$  has the form  $\mu \mapsto \langle \mu, f \rangle$  for some  $f \in \mathcal{F}$ , which explains why  $\tilde{\Phi}'(\mu_0)$  lives in  $\mathcal{F}$  and not in  $\mathcal{F}^{**} \setminus \mathcal{F}$ . Besides it is known [22] that  $\tilde{\Phi}'(\mu_0) \in \mathcal{F}$  is determined by (2) uniquely up to constants on  $U$ .

If  $\mu_0$  solves (RP) and  $\tilde{\Phi}$  has the Gâteaux differential at  $\mu_0$ , then it must belong to the negative normal cone to  $M(\mathcal{F})$ ; this means

$$(3) \quad \tilde{\Phi}'(\mu_0) \in -N_{M(\mathcal{F})}(\mu_0),$$

where  $N_{M(\mathcal{F})}(\mu_0)$  denotes the normal cone  $\{f \in \mathcal{F} : \forall \mu \in M(\mathcal{F}) : \langle \mu - \mu_0, f \rangle \leq 0\}$ . If  $N_{M(\mathcal{F})}(\mu_0) = \{\text{constants}\}$ , then (3) turns out to the condition  $\tilde{\Phi}'(\mu_0)$  to be

constant, which is essentially the standard Euler-Lagrange equation  $\bar{\Phi}'(\mu_0) = 0$  because the constants are always included in  $N_{M(\mathcal{F})}(\mu_0)$  so that they are factually irrelevant.

However, in nonlinear optimal control problems where a finer relaxation is generally inevitable the normal cone  $N_{M(\mathcal{F})}(\mu_0)$  is typically much larger, cf. also Section 4 below.

In view of the density of  $c(U)$  in  $M(\mathcal{F})$ , we can obviously rewrite (2) (3) as

$$(4) \quad \langle \mu_0, \bar{\Phi}'(\mu_0) \rangle = \sup_{u \in U} [\bar{\Phi}'(\mu_0)](u).$$

We will refer to (4) as an *abstract maximum principle*.

Let us only outline the situation when (P) comes from an optimal control problem having an abstract structure

$$(OCP) \quad \begin{cases} \text{Minimize } \varphi(u, y) \\ \text{subject to } u \in U, y \in Y, A(u, y) = 0, \end{cases}$$

where  $\varphi : U \times Y \rightarrow \mathbb{R}$  is a cost function,  $Y$  a Banach space of states, and  $A : U \times Y \rightarrow \Lambda$  with  $\Lambda$  another Banach space determines a state operator  $\pi : U \rightarrow Y$  such that  $y = \pi(u)$  if and only if  $A(u, y) = 0$ . We will suppose that the convex compactification  $M(\mathcal{F})$  of  $U$  is so fine that both  $A$  and  $\varphi$  admit continuous extensions  $\bar{\varphi} : M(\mathcal{F}) \times Y \rightarrow \mathbb{R}$  and  $\bar{A} : M(\mathcal{F}) \times Y \rightarrow \Lambda$  and, moreover, for every  $\mu \in M(\mathcal{F})$  the equation  $\bar{A}(\mu, y) = 0$  has a unique solution  $y = \bar{\pi}(\mu)$  and the state operator  $\bar{\pi} : M(\mathcal{F}) \rightarrow Y$  thus determined is continuous (then  $\bar{\pi}$  is the continuous extension of the original operator  $\pi : U \rightarrow Y$  in the sense that  $\bar{\pi} \circ c = \pi$ ). We can then define the following relaxed optimal control problem:

$$(ROCP) \quad \begin{cases} \text{Minimize } \bar{\varphi}(\mu, y) \\ \text{subject to } \mu \in M(\mathcal{F}), y \in Y, \bar{A}(\mu, y) = 0. \end{cases}$$

The existence of a solution (ROCP) (considered as a generalized solution of (OCP)) and stability of the set of all these solutions is again ensured by the standard compactness and continuity arguments.

The problem (ROCP) can be equivalently written in the form of (RP) by putting  $\bar{\Phi}(\mu) = \bar{\varphi}(\mu, \bar{\pi}(\mu))$ . Let us suppose that  $\bar{\Phi} : M(\mathcal{F}) \rightarrow \mathbb{R}$  is Gâteaux differentiable. By using the adjoint-equation technique developed essentially in [16] we can evaluate the Gâteaux differential  $\bar{\Phi}'(\mu_0)$  as follows: Let us assume that  $\bar{\varphi}(\mu_0, \cdot)$  has the continuous Fréchet derivative  $\bar{\varphi}'_{\mu_0} : M(\mathcal{F}) \times Y \rightarrow Y^*$  and  $\bar{\varphi}(\cdot, y)$  has a Gâteaux derivative  $\bar{\varphi}'_y \in \mathcal{F}$ , and the same holds for  $\bar{A}$  weakly; this means there is  $\bar{A}'_y : M(\mathcal{F}) \times Y \rightarrow \mathcal{L}(Y, \Lambda)$  such that  $\langle \lambda, \bar{A}'_y(\mu_0, y) \rangle \in Y^*$  is the Fréchet derivative of  $\langle \lambda, \bar{A}(\mu_0, \cdot) \rangle : Y \rightarrow \mathbb{R}$  at  $y \in Y$  and  $\langle \lambda, \bar{A}'_y \rangle : M(\mathcal{F}) \times Y \rightarrow Y^*$  is continuous for all  $\lambda \in \Lambda^*$ , and there is  $\bar{A}'_{\mu_0} : M(\mathcal{F}) \times Y \rightarrow \mathcal{L}(\mathcal{F}, \Lambda)$  such that  $\langle \lambda, \bar{A}'_{\mu_0}(\mu_0, y) \rangle \in \mathcal{F}$  is the Gâteaux derivative of  $\langle \lambda, \bar{A}(\cdot, y) \rangle$  at  $\mu_0 \in M(\mathcal{F})$  for all  $\lambda$ ;  $\mathcal{L}(\cdot, \cdot)$  denotes the space of all continuous, linear operators. Moreover, let the state operator  $\bar{\pi} : M(\mathcal{F}) \rightarrow Y$

be directionally Lipschitz continuous, i.e.  $\|\pi(\mu_0 + h(\mu - \mu_0)) - \pi(\mu_0)\| \leq L_{\mu_0, \mu} h$ . If the adjoint equation

$$(5) \quad [\bar{A}'_y(\mu_0, \bar{\pi}(\mu_0))]^* \lambda + \bar{\varphi}'_y(\mu_0, \bar{\pi}(\mu_0)) = 0$$

has a solution  $\lambda \in \Lambda^*$  (=the so-called adjoint state), then we can evaluate  $\Phi'(\mu_0) = \bar{\varphi}'_u(\mu_0, \bar{\pi}(\mu_0)) + [\bar{A}'_u(\mu_0, \bar{\pi}(\mu_0))]^* \lambda + \text{const.}$  Obviously, (4) is then equivalent to (5) completed with the maximum principle

$$(6) \quad \langle \mu_0, f_{\lambda, y, \mu_0} \rangle = \sup_{u \in U} f_{\lambda, y, \mu_0}(u),$$

where  $f_{\lambda, y, \mu_0} \in \mathcal{F}$  is an "abstract Hamiltonian" defined by

$$(7) \quad f_{\lambda, y, \mu_0} = \bar{\varphi}'_u(\mu_0, y) + [\bar{A}'_u(\mu_0, y)]^* \lambda + \text{const.} \quad \text{with } y = \pi(\mu_0).$$

We will see in Section 4 that in concrete relaxed problems the "Hamiltonian" from (7) actually gives the standard Hamiltonians which typically do not depend explicitly on  $\mu_0$  provided  $\mathcal{F}$  is large enough so that both  $\varphi(\cdot, y) \in \mathcal{F}$  and  $\langle \lambda, A(\cdot, y) \rangle \in \mathcal{F}$  for any  $\lambda \in \Lambda$ .

### 3. Concrete convex compactifications of $U$ .

In applications, relaxed controls are typically the so-called Young measures (i.e. parameterized probability measures) or some generalizations of them. It is related with a concrete form of the set of controls  $U$ . For simplicity, let us consider here only bounded controls in the form

$$(8) \quad U = \{u \in L^\infty(Q; \mathbb{R}^k); \forall \xi \in Q: u(\xi) \in S\}$$

with  $Q \subset \mathbb{R}^n$  and  $S \subset \mathbb{R}^k$  compact. Such  $U$  appears quite typically in optimal control of distributed parameter systems. Besides, let us suppose that  $n = n_P + n_R$  and  $Q$  admits a decomposition  $Q = P \times R$ , with  $P \subset \mathbb{R}^{n_P}$  and  $R \subset \mathbb{R}^{n_R}$ ,  $n_P \geq 1$ ,  $n_R \geq 0$ . In the case  $n_R = 0$ , we put simply  $Q = P$ . Roughly speaking,  $P$  denotes the "directions" in  $Q$  where the "rapid" oscillations of the controls appear. Let us put

$$(9) \quad B_R = \{R \rightarrow S \text{ measurable}\},$$

which is a bounded subset of  $L^\infty(R; \mathbb{R}^k)$ ; for  $n_R = 0$  we put simply  $B_R = S \subset \mathbb{R}^k$ . Let us choose a linear subspace  $H \subset L^1(P; C(B_R))$  and define the mapping  $\Psi_P: H \rightarrow C(U)$  by  $\Psi_P(h) = (u \mapsto \int_P h(\zeta, u(\zeta, \cdot)) d\zeta)$ ; note that  $\Psi_P(h)$  is nothing else than the integral over  $P$  of the Nemitskii operator generated by  $h$ . Furthermore, let us define the imbedding  $i_H: U \rightarrow H^*: u \mapsto [\Psi_P(\cdot)](u)$  and denote by  $Y_H(P; B_R)$  the weak\* closure of  $i_H(U)$  in  $H^*$ .

**Lemma 1.** (Cf. [20, 24]) Let  $H$  be a linear subspace of  $L^1(P; C(B_R))$  and  $\mathcal{F} = \mathcal{F}_H = \Psi_P(H) + \text{const.}$  Then  $\mathcal{F}_H$  is a convexifying subspace of  $C(U)$  with  $U$  from

(8), and the convex compactification  $(M(\mathcal{F}_H), \mathcal{F}_H^*, c)$  of  $U$  is equivalent with  $(Y_H(P; B_R), H^*, i_H)$ . More precisely, the affine homeomorphism  $M(\mathcal{F}_H) \xrightarrow{\sim} Y_H(P; B_R)$  fixing  $U$  is the adjoint operator  $\Psi_P^* : \mathcal{F}^* \rightarrow H^*$  restricted to  $M(\mathcal{F})$ .

Supposing that  $C(P) \cdot H = H$  (this means  $g \in C(P)$  and  $h \in H$  implies  $g \cdot h \in H$  for  $[g \cdot h](\zeta, \cdot) = g(\zeta) \cdot h(\zeta, \cdot)$ ), we can define the bilinear mapping  $\bullet : H \times Y_H(P; B_R) \rightarrow L^1(P)$  by  $h \bullet \eta = \text{w-lim}_\alpha h \circ u_\alpha$  for  $i_H(u_\alpha) \rightarrow \eta$ , where  $[h \circ u](\zeta) = h(\zeta, u(\zeta, \cdot))$ . Note that the net  $\{|h \circ u_\alpha|\}$  has an integrable majorant so that its weak limit in  $L^1(\Omega)$  does exist at least if a subnet is taken, but it exists even for the whole net because this limit is unique and equals  $h \bullet \nu \in L^1(\Omega)$  defined alternatively by  $\langle h \bullet \nu, g \rangle = \langle \eta, g \cdot h \rangle$  for every  $g \in C(P)$ .

Let us come back to Sect. 2 and suppose that  $\Phi : M(\mathcal{F}_H) \rightarrow \mathbb{R}$  is Gâteaux differentiable at  $\mu_0$ . By Lemma 1, we can transfer our considerations to  $Y_H(P; B_R)$ : Let us put  $\tilde{\Phi} = \Phi \circ (\Psi_P^*)^{-1} : Y_H(P; B_R) \rightarrow \mathbb{R}$ . Then also  $\tilde{\Phi}$  is Gâteaux differentiable at  $\eta_0 = \Psi_P^* \mu_0 \in Y_H(P; B_R)$  and  $\tilde{\Phi}'(\eta_0) \in H$  is determined by  $\Psi_P^{-1}(\tilde{\Phi}'(\mu_0)) + h_0$  with arbitrary  $h_0 \in H$  such that  $h_0(\zeta, \cdot)$  is constant on  $B_R$  for a.a.  $\zeta \in P$ . We will call  $\mathcal{H} = \tilde{\Phi}'(\eta_0) \in H$  the "Hamiltonian"; obviously,  $\mathcal{H}$  is determined uniquely up to  $H \cap [L^1(P) \odot 1_R]$  where  $1_R$  denotes the function  $R \rightarrow \{1\}$ .

**Theorem 2.** The abstract maximum principle (4) is equivalent

$$(10) \quad \int_P [\mathcal{H} \bullet \eta_0](\zeta) d\zeta = \sup_{u \in U} \int_P \mathcal{H}(\zeta, u(\zeta, \cdot)) d\zeta.$$

If  $H \subset L^1(P; C(B_R^\tau))$  where  $B_R^\tau$  denotes the set  $B_R$  endowed by a topology  $\tau$  which makes it compact or a complete metrizable separable space, then (10) is also equivalent with

$$(11) \quad [\mathcal{H} \bullet \eta_0](\zeta) = \sup_{v \in B_R} \mathcal{H}(\zeta, v) \quad \text{for a.a. } \zeta \in P.$$

*Sketch of the proof.* As for the equivalence of (4) with (10), it suffices to realize the identity  $\int_P [\mathcal{H} \bullet \eta](\zeta) d\zeta = \langle \eta, \mathcal{H} \rangle = \langle \Psi_P^* \mu, \mathcal{H} \rangle = \langle \mu, \Psi_P \mathcal{H} \rangle = \langle \mu, \Phi'(\mu_0) \rangle$  for any  $\eta = \Psi_P^* \mu \in Y_H(P; B_R)$ .

As for the localization of (10) to (11), we only need to make a measurable selection from a multivalued mapping of the type  $\zeta \mapsto \{v \in B_R; \mathcal{H}(\zeta, v) \geq c(\zeta)\}$  with a suitable  $c \in L^1(P)$ , which needs the assumptions on  $\tau$  as imposed above; for details we refer to [23, Theorem 2].  $\square$

In accord with usual terminology, we will call (10) and (11) the integral and the Pontryagin maximum principles, respectively; cf. [14, 19]. Let us point out that the Hamiltonian  $\mathcal{H}$  (determined uniquely only up to integrands  $c \in H$  such that  $c(\zeta, \cdot)$  is constant) resulting from our theory coincides for the choice  $c = 0$  with the usual "guessed" Hamiltonian; cf. [23, Sect. 4] for the case of a finite-dimensional



system or (21) below for the parabolic system. Of course, the constancy (on  $P$ ) of the Hamiltonian along optimal trajectories, which is usually included into the Pontryagin principle, is irrelevant here because our Hamiltonian is unique only up to integrable functions on  $P$ . On the other hand, the Hamiltonian has a definite meaning by our derivation, namely it represents the Gâteaux differential with respect to the geometry coming from  $H^*$ .

After the transformation of a particular  $H \subset L^1(P; C(B_R))$  via  $\Psi_P$ , we get the subspace  $\mathcal{F}_H$  of  $C(U)$  without any explicit reference to the decomposition  $Q = P \times R$ , which makes possible a comparison between various  $H$  even if the decomposition  $Q = P \times R$  vary. The following assertion is a consequence of Theorem 1 with Lemma 1. Note that it identifies the topology on  $L^1(P; C(B_R))$  induced projectively via  $\Psi_P$  from  $C(U)$  as the decisive topology in the sense that the closure of  $H$  in this topology does not change the resulting convex compactification but any further enlargement does refine it.

**Theorem 3.** *Let  $Q = P_1 \times R_1$  and  $Q = P_2 \times R_2$  be two admissible decompositions of  $Q$  and let the linear subspaces  $H_1 \subset L^1(P_1; C(B_{R_1}))$  and  $H_2 \subset L^1(P_2; C(B_{R_2}))$  be given such that  $\mathcal{F}_{H_1} \subset \mathcal{F}_{H_2}$ . Then  $M(\mathcal{F}_{H_1}) \preceq M(\mathcal{F}_{H_2})$ , which means just that  $Y_{H_1}(P_1, B_{R_1})$  is a coarser convex compactification of  $U$  than  $Y_{H_2}(P_2, B_{R_2})$ . Moreover,  $cl_{C(U)} \mathcal{F}_{H_1} = cl_{C(U)} \mathcal{F}_{H_2}$  if and only if  $M(\mathcal{F}_{H_1}) \cong M(\mathcal{F}_{H_2})$ .*

We can see that there is, in fact, a large freedom in the choice of the decomposition  $Q = P \times R$  and then in the choice of the particular subspace  $H \subset L^1(P; C(B_R))$ . It is obvious that the choice of the decomposition of  $Q$  has an immediate impact on the character of the Pontryagin maximum principle (11). Namely, the larger the component  $P$  (which contains directions in which "rapid" oscillations in controls are allowed), the more local the Pontryagin maximum principle that can be obtained. From this point of view, the best choice of the decomposition is  $Q = P$ ,  $n = n_P$ ,  $n_R = 0$ . On the other hand, another decomposition with  $n_R > 0$  can yield finer convex compactifications, which might be sometimes inevitable when special problems are to be treated; cf. Sect. 4. As far as the choice of  $H$  is concerned, we can say that a larger  $H$  makes the resulting convex compactification finer (cf. also Theorem 3) and thus the class of mappings which admit a continuous extension to this compactification is larger. Also we can say that a "limit" information about oscillations of  $\varepsilon$ -optimal controls contained in the relaxed optimal controls is then greater but, of course, the implementation of such more informative solutions on computers is harder. Taking  $H$  smaller has naturally just opposite effects. This is summarized in the following tables.

	$H$	
	larger	smaller
convex compactification	finer	coarser
information in a solution	greater	lesser
implementation	harder	easier
a class of continuously extendable mappings	more	less

TABLE 1. The influence of the choice of the linear subspace  $H$ .

	$P$	
	larger	smaller
optimality conditions	more local	less local
directions for "rapid" oscillations in controls	wider	narrower

TABLE 2. The influence of the choice of the decomposition  $Q = P \times R$ .

Given a class of problems to be relaxed, an attempt for a general recommendation might look like:  $n_P$  and  $P$  should be chosen as large as possible to localize the maximum principle as much as possible, and simultaneously  $H$  should be taken as small as possible (with respect to a given class of problems that are to admit a continuous extension on the resulted convex compactification) to get a convex compactification as coarse as possible (which makes its implementation easier). Also, unnecessarily finer convex compactifications deteriorate a chance that, at least for some special problems from a given class, the solution of the relaxed problem is unique. Therefore, minimal convex compactifications satisfying some prescribed conditions are of a particular interest, cf. also [21, Theorem 3].

#### 4. Examples.

We want to deal with the particular case  $Q = (0, T) \times \Omega$  with  $\Omega$  a domain in  $\mathbb{R}^m$ . It appears typically in distributed control of evolution partial differential equations, say parabolic or hyperbolic. (Quite equally we could treat the case  $Q = (0, T) \times \partial\Omega$  with  $\partial\Omega$  an  $(m-1)$ -dimensional boundary of  $\Omega$ , which appears in a boundary control of such equations.) Then time  $t$  ranges the interval  $(0, T)$  while  $x$  ranges the spatial domain  $\Omega$ . In the previous notation,  $\xi = (t, x)$ , so that the set of controls, cf. (8), is now

$$(12) \quad U = \{u \in L^\infty((0, T) \times \Omega; \mathbb{R}^k); u(t, x) \in S \text{ for a.a. } t \in (0, T), x \in \Omega\}.$$

There are now two reasonable decompositions of  $Q \equiv (0, T) \times \Omega = P \times R$ , namely

1.  $P = (0, T) \times \Omega$ ,  $n = n_P = m + 1$ ,  $n_R = 0$ ; then  $B_R = S$ ,

2.  $P = (0, T)$ ,  $R = \Omega$ ,  $n_P = 1$ ,  $n_R = m$ ; then  $B_R \equiv B_\Omega = \{u \in L^\infty(\Omega; \mathbb{R}^k); u(x) \in S \text{ for a.a. } x \in \Omega\}$ .

Of course, we could also imagine other decompositions of  $Q = (0, T) \times \Omega$  but the corresponding relaxations do not seem to be reported in the literature.

Let us start with the first possibility. The resulting Pontryagin maximum principle is then formulated for a.a.  $t \in (0, T)$  and a.a.  $x \in \Omega$ , and the supremum in (11) is taken over  $S$ . Such kind of maximum principles has been treated by Ahmed, Teo [3], Sadigh-Esfandiari, Sloss, Bruch Jr., Sadek [25, 26], Zolezzi [27], etc.

Though there are uncountably many convex compactifications of the form  $M(\mathcal{F}_H)$  with  $H \subset L^1((0, T) \times \Omega; C(S))$ , we will mention only two of them as examples. The first one is the finest convex compactification in this class, which takes

$$(13) \quad H = L^1((0, T) \times \Omega; C(S)) ,$$

while the second one uses

$$(14) \quad H = L^1((0, T) \times \Omega; \mathcal{L}(S)) ,$$

where  $\mathcal{L}(S)$  denotes the set of all linear functions  $\mathbb{R}^k \rightarrow \mathbb{R}$  restricted to  $S$ ; in other words,  $H$  from (14) contains just the functions of the form  $(t, x, s) \mapsto \sum_{i=1}^k g_i(t, x) s_i$  with  $g_i \in L^1((0, T) \times \Omega)$  and  $s = (s_1, \dots, s_k) \in S \subset \mathbb{R}^k$ .

In the case (13), it is known that  $H^* = L^1((0, T) \times \Omega; C(S))^*$  is isometrically isomorphic with  $L_w^\infty((0, T) \times \Omega; rca(S))$  via the mapping  $\psi : L_w^\infty((0, T) \times \Omega; rca(S)) \rightarrow L^1((0, T) \times \Omega; C(S))^*$  defined by

$$(15) \quad \nu = \{\nu_{t,x}\}_{t \in (0,T), x \in \Omega} \mapsto \left( h \mapsto \int_0^T \int_\Omega \int_S h(t, x, s) \nu_{t,x}(ds) dx dt \right)$$

where  $L_w^\infty((0, T) \times \Omega; rca(S))$  is the space of weakly\* measurable essentially bounded mappings from  $(0, T) \times \Omega$  to the space  $rca(S)$  of regular bounded  $\sigma$ -additive set functions (=Borel measures) on  $S$ . According to the usual notation, we wrote  $\nu_{t,x}$  instead of  $\nu(t, x)$ , which is to emphasize that  $\nu$  is a parameterized measure. Moreover,  $\psi^{-1}$  maps  $Y_H((0, T) \times \Omega; S)$  onto

$$\begin{aligned} \mathcal{Y}((0, T) \times \Omega; S) &= \{ \nu \in L_w^\infty((0, T) \times \Omega; rca(S)); \nu_{t,x} \in rca_+^1(S) \\ &\quad \text{for a.a. } t \in (0, T), x \in \Omega \} , \end{aligned}$$

where  $rca_+^1(S) = \{ \nu \in rca(S); \nu \geq 0 \text{ \& } \nu(S) = 1 \}$  denotes the set of all regular probability measures of  $S$ . The elements of  $\mathcal{Y}((0, T) \times \Omega; S)$  are just the so-called Young measures, parameterized by  $(t, x) \in (0, T) \times \Omega$  and supported on  $S$ .

In the case (14), it is easy to show that  $Y_H((0, T) \times \Omega; S)$  is affinely homeomorphic with

$$\{ u \in L^\infty((0, T) \times \Omega; \mathbb{R}^k); u(t, x) \in \overline{\text{co}}(S) \text{ for a.a. } t \in (0, T), x \in \Omega \}$$

endowed with the weak\* topology of  $L^\infty((0, T) \times \Omega; \mathbb{R}^k)$ . The homeomorphism is via the mapping which assigns each  $u \in L^\infty((0, T) \times \Omega; \mathbb{R}^k)$  the linear functional  $[\Psi_Q(\cdot)](u) \in H^*$ . Note that this mapping is linear since all the integrands from  $H$  from (14) are linear in terms of  $s \in \mathbb{R}^k$ . This gives basically the standard relaxation which uses more or less the original controls equipped with the  $L^\infty$ -weak\* topology. However, such coarse relaxation is suitable only for problems having a convex/linear structure, possibly with slightly nonlinear perturbations.

Let us go on to the second mentioned decomposition of  $(0, T) \times \Omega$ , namely  $P = (0, T)$  and  $R = \Omega$ . The resulting Pontryagin maximum principle is then formulated for a.a.  $t \in (0, T)$ , and the supremum in (11) is taken over  $B_\Omega = \{u \in L^\infty(\Omega; \mathbb{R}^k); u(x) \in S \text{ for a.a. } x \in \Omega\}$ . Such kind of relaxations and/or maximum principles has been treated by Ahmed [1, 2], Basile, Mininni [4], Fattorini [10], Frankowska [12, 13], Kampowsky [15], Papageorgiou [17], etc.

Again there are uncountably many convex compactifications of the form  $M(\mathcal{F}_H)$  with  $H \subset L^1(0, T; C(B_\Omega^\tau))$  with  $\tau$  some Hausdorff topology on  $B_\Omega$ , but we will mention only two of them as examples. The first one is

$$(16) \quad H = L^1(0, T; C(B_\Omega^\tau)) \quad , \quad \tau = \text{the norm topology from } L^p(\Omega; \mathbb{R}^k) \quad ,$$

while the second one uses

$$(17) \quad H = L^1(0, T; C(B_\Omega^\tau)) \quad , \quad \tau = \text{the weak* topology from } L^\infty(\Omega; \mathbb{R}^k) \quad .$$

Note that both (16) for  $1 \leq p < +\infty$  and (17) for  $S \subset \mathbb{R}^k$  convex satisfy the assumptions of Theorem 2 needed to establish the Pontryagin maximum principle.

As to the case (16),  $B_\Omega^\tau$  is not compact and  $H^* = L^1(0, T; C(B_\Omega^\tau))^*$  is isometrically isomorphic (again via  $\psi$  defined like in (15)) with  $L_w^\infty(0, T; rba(B_\Omega^\tau))$ , where  $rba(B) \cong C(B)^*$  denotes the set of all regular bounded additive set functions on a normal topological space  $B$ . Then elements of  $Y_H(0, T; B_\Omega^\tau)$  can be interpreted as probability regular bounded additive set functions, that means elements of  $L_w^\infty(0, T; rba(B_\Omega^\tau))$  valued in  $rba_1^+(B_\Omega^\tau)$ . In the context of optimal-control theory, elements of  $rba(B)$  are also called "finite additive measures" to distinguish it from the usual measures which are  $\sigma$ -additive. Finitely additive measures were used in this context in the works by Fattorini [8, 9, 10, 11], and in general optimization theory also by Chentsov [6, 7], Pashaev [18], etc. Let us only mention that, since  $rba(B_\Omega^\tau) \cong rca(B_\Omega^\tau)$  with  $B_\Omega^\tau$  denoting the Stone-Ćech compactification of  $B_\Omega^\tau$ , the finitely additive measures on a normal space  $B_\Omega^\tau$  can be equally understood as standard  $\sigma$ -additive measures of a (non-metrizable) compact space  $B_\Omega^\tau$ .

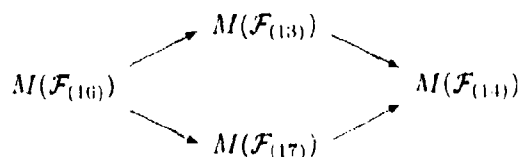
Let us suppose, for simplicity, that  $S \subset \mathbb{R}^k$  is convex. Since  $B_\Omega^\tau$  is compact provided  $\tau$  is the weak\* topology of  $L^\infty(\Omega; \mathbb{R}^k)$ , it is now evident that  $H^*$  with  $H$  from (17) is isometrically isomorphic with  $L_w^\infty(0, T; rca(B_\Omega^\tau))$ , and  $Y_H(0, T; B_\Omega^\tau)$  is mapped by this isomorphism onto

$$\mathcal{Y}(0, T; B_\Omega^\tau) = \{\nu \in L_w^\infty(0, T; rca(B_\Omega^\tau)); \nu_t \in rca_1^+(B_\Omega^\tau) \text{ for a.a. } t \in (0, T)\} \quad .$$

The elements of  $\mathcal{Y}((0, T) \times \Omega; S)$  are just the Young measures, parameterized by  $t \in (0, T)$  and supported on  $B_\Omega^S$ . Such Young measures has been basically used by Papageorgiou [17].

The relations between the mentioned concrete convex compactifications of  $U$  are stated in the following assertion. For brevity, we will denote by  $H_{(\cdot)}$  and  $\mathcal{F}_{(\cdot)}$  respectively the subspace  $H$  from  $(\cdot)$  and  $\mathcal{F}_H \subset C(U)$  created by  $H_{(\cdot)}$ ; for example  $\mathcal{F}_{(16)}$  stands for  $\mathcal{F}_H$  with  $H$  from (16), denoted by  $H_{(16)}$ .

**Theorem 4.** *The relations between the convex compactifications  $M(\mathcal{F}_{(13)})$ ,  $M(\mathcal{F}_{(14)})$ ,  $M(\mathcal{F}_{(16)})$ , and  $M(\mathcal{F}_{(17)})$  are given by the following diagram*



where each arrow indicates the existence of an affine homeomorphism; in other words, each arrow goes from a finer convex compactification to a coarser one. Besides, no arrow can be reversed if  $S$  is not a singleton; this means the original convex compactification is strictly finer than the terminal one. Finally, the convex compactifications  $M(\mathcal{F}_{(13)})$  and  $M(\mathcal{F}_{(17)})$  are actually not comparable.

*Sketch of the proof.* The fact that  $M(\mathcal{F}_{(13)}) \supseteq M(\mathcal{F}_{(14)})$  follows immediately from  $H_{(13)} \supset H_{(14)}$ . Taking two points  $s_1, s_2 \in S$  and two sequences  $\{u_k^1\}$  and  $\{u_k^2\}$ , the former one taking the values  $s_1$  and  $s_2$  and converging weakly\* to the constant  $\frac{1}{2}s_1 + \frac{1}{2}s_2 \in S$  while the latter one being constant  $u_k^2 \equiv \frac{1}{2}s_1 + \frac{1}{2}s_2$ . Now it is clear that the limit of  $\{c(u_k^j)\}_{k \in \mathbb{N}}$  in  $M(\mathcal{F}_{(14)})$  exists and is the same for  $j = 1$  and  $2$ , while these sequences can be separated on  $M(\mathcal{F}_{(13)})$  by, for example, a function  $f = \Psi_Q(1_Q \odot v)$  with  $v \in C(S)$  such that  $v(s_1) = v(s_2) = 0$  and  $v(\frac{1}{2}s_1 + \frac{1}{2}s_2) = 1$  because obviously  $\langle c(u_k^1), f \rangle = \int_Q v(u_k^1(t, x)) dt dx = 0$  while  $\langle c(u_k^2), f \rangle = |Q| > 0$  for any  $k \in \mathbb{N}$ . This shows that  $M(\mathcal{F}_{(13)}) \supsetneq M(\mathcal{F}_{(14)})$ .

By analogous arguments we can show also  $M(\mathcal{F}_{(16)}) \supsetneq M(\mathcal{F}_{(17)})$ . As for the two sequences which have the same limits on  $M(\mathcal{F}_{(17)})$  but can be separated on  $M(\mathcal{F}_{(16)})$ , we can take  $u_k^1(t, x) = v_k(x)$  and  $u_k^2$  as previously, where  $v_k : \Omega \rightarrow S$  converges to  $\frac{1}{2}s_1 + \frac{1}{2}s_2$  weakly\* in  $L^\infty(\Omega; \mathbb{R}^k)$  but not strongly in  $L^p(\Omega; \mathbb{R}^k)$ .

As for  $M(\mathcal{F}_{(17)}) \supseteq M(\mathcal{F}_{(14)})$ , we cannot show directly  $H_{(17)} \supset H_{(14)}$  because the partition  $Q = P \times R$  is not the same. Nevertheless, we can show that  $\mathcal{F}_{(17)} \supset \mathcal{F}_{(14)}$  because for any  $\sum_{l=1}^k g_l \odot s_l \in H_{(14)}$  there is  $h \in H_{(17)}$  such that  $\Psi_Q(\sum_{l=1}^k g_l \odot s_l) = \Psi_{(0,T)}(h)$ , namely  $h(t, v) = \int_\Omega \sum_{l=1}^k g_l(t, x) v_l(x) dx$  for  $t \in (0, T)$  and  $v \in B_\Omega$ ; note that the following estimate holds:  $\|h\|_{L^1((0,T);C(B_\Omega^S))} \leq \sup_{s \in S} |s| \sum_{l=1}^k \|g_l\|_{L^1((0,T) \times \Omega)}$ .

Analogous arguments can be also used for  $M(\mathcal{F}_{(16)}) \succ M(\mathcal{F}_{(13)})$ . Indeed, having  $h \in H_{(13)}$  we can find  $h_1 \in H_{(16)}$  such that  $\Psi_{(0,T)}(h_1) = \Psi_Q(h)$ , namely  $h_1(t, v) = \int_{\Omega} h(t, x, v(x)) dx$ . Note that the continuity of  $h_1(t, \cdot) : B_{\Omega}^T \rightarrow \mathbb{R}$  follows by the standard properties of the Nemytskii mappings, while the estimate  $\|h_1\|_{L^1(0,T;C(B_{\Omega}^T))} \leq \|h\|_{L^1((0,T) \times \Omega; C(S))}$  is obvious. As for the two sequences, let us take  $\{u_k^1\}$  as in the first case but with  $v_k^1(t, \cdot)$  constant (and equal to either  $s_1$  or  $s_2$ ) on  $\Omega$ , and  $\{u_k^2\}$  such that  $u_k^2(t, x) = u_k^1(t, x)$  if  $x \in \Omega'$  and  $u_k^2(t, x) = s_2$  if  $u_k^1(t, x) = s_1$  and  $u_k^2(t, x) = s_1$  if  $u_k^1(t, x) = s_2$  provided  $x \in \Omega''$ , where  $\Omega'$  and  $\Omega''$  are disjoint parts of  $\Omega$  of positive measure. Roughly speaking, both sequences converges to the Young measure  $\nu \in L_w^{\infty}(Q; \text{rea}(S))$  given by  $\nu_{t,x} = \frac{1}{2}\delta_{s_1} + \frac{1}{2}\delta_{s_2}$  in the representation of  $M(\mathcal{F}_{(13)})$ . On the other hand, on the representation of  $M(\mathcal{F}_{(16)})$ , the first one converges to the Young measure  $\nu^1 \in L_w^{\infty}(0, T; \text{rba}(B_{\Omega}^T))$  given by  $\nu_t = \frac{1}{2}\delta_{v_1^1} + \frac{1}{2}\delta_{v_2^1}$  with  $v_j^1(x) = s_j$  for  $j = 1, 2$  while the second one converges to  $\nu^2 \in L_w^{\infty}(0, T; \text{rba}(B_{\Omega}^T))$  given by  $\nu_t = \frac{1}{2}\delta_{v_1^2} + \frac{1}{2}\delta_{v_2^2}$  with  $v_1^2(x) = s_1$  if  $x \in \Omega'$  and  $v_1^2(x) = s_2$  if  $x \in \Omega''$ , and  $v_2^2(x) = s_2$  if  $x \in \Omega'$  and  $v_2^2(x) = s_1$  if  $x \in \Omega''$ .

By this construction we can also show that  $M(\mathcal{F}_{(17)}) \not\preceq M(\mathcal{F}_{(13)})$ , while the fact that  $M(\mathcal{F}_{(17)}) \not\preceq M(\mathcal{F}_{(13)})$  can be proved similarly as the fact that  $M(\mathcal{F}_{(16)}) \not\preceq M(\mathcal{F}_{(17)})$ .  $\square$

Finally, let us illustrate the above theory by one model example of an optimal control of a nonlinear parabolic equation. We will take (OCP) from Section 2 with the following data:  $k = 1$ ,  $U$  from (12) with  $S = [s_1, s_2]$ ,  $Y = H^1(Q) \cap L^2(0, T; H_0^1(\Omega))$ ,  $\Lambda = L^2(0, T; H^{-1}(\Omega)) \times L^2(\Omega)$ , the cost function

$$(18) \quad \varphi(u, y) = \int_0^T \int_{\Omega} \left( u(t, x) - \frac{1}{|\Omega|} \int_{\Omega} u(t, \zeta) d\zeta \right)^2 dx dt + \int_{\Omega} (y(T, x) - y_d(x))^2 dx,$$

with the desired terminal profile  $y_d \in L^2(\Omega)$ , and the state equation  $A(y, u) = 0$  in the form

$$(19) \quad \frac{\partial y}{\partial t} - \Delta y + \gamma(y) = \gamma(u), \quad y(0, \cdot) = y_0,$$

with  $\gamma : \mathbb{R} \rightarrow \mathbb{R}$  smooth and nondecreasing, the initial data  $y_0 \in H_0^1(\Omega)$  and, for simplicity, the Dirichlet boundary conditions  $y(t, \cdot) = 0$  on  $\partial\Omega$ . More precisely, the operator  $A : U \times Y \rightarrow \Lambda$  is determined by the standard weak formulation of (19) with the boundary and initial conditions, which means that  $A(u, y) = (z_1, z_0) \in L^2(0, T; H^{-1}(\Omega)) \times L^2(\Omega)$  with  $z_0 = y(T, \cdot) - y_d$  and  $z_1 = \partial y / \partial t - \Delta y + \gamma(y) - \gamma(u)$  understood in the distributional sense. Note that  $A(u, y) = 0$  has actually the only weak solution  $y = \pi(u) \in H^1(Q) \cap L^2(0, T; H_0^1(\Omega))$ . The growth of the nonlinearity  $\gamma$  at infinity is irrelevant because  $L^{\infty}$ -estimates are at our disposal thanks to the comparison principle.

The interpretation of the problem (OCP) with (18)–(19) might be the following: the optimal control should drive the parabolic system from a given initial

state  $y_0$  as close to the desired terminal state  $y_d$  as possible and, simultaneously, the spatial profiles of the control should be "as constant as possible". For example, if (19) has a heat-transfer interpretation (i.e.  $y$  is a temperature) and  $\gamma(y) = y^4$ , then we can say that the control acts in (19) via a Stefan-Boltzmann radiation term (which is of the 4th power).

It is obvious that, supposing  $\gamma$  nonlinear, such (OCP) does not bear a relaxation by means of  $H$  from (14) because (19) could not be extended continuously. For the same reason, also (17) is unsatisfactory. As for the choice (13), it would enable a continuous extension of (19) but not of (18). This can be seen easily by taking the two sequences constructed in the proof of Theorem 4 (the part  $M(\mathcal{F}_{(16)}) \succ M(\mathcal{F}_{(13)})$ ). On the other hand, the choice (16) is satisfactory both for (18) and (19) provided  $p \geq 2$ . Thus we demonstrated a nontrivial situation when we must inevitably choose the finest convex compactification from those investigated by Theorem 4. Let us also remark that the transformation of control variable  $\gamma(u) \rightarrow \tilde{u}$  would not change the situation because then the failure of the weak continuity, that appears on coarser convex compactifications, would be only transferred from (19) to (18) provided  $\gamma$  is actually nonlinear.

If one evaluates (5) in our case, one gets after integration by parts (cf. [23] for this procedure) the equations for the adjoint state  $\lambda = (\lambda_1, \lambda_0) \in \Lambda^* = L^2(0, T; H_0^1(\Omega)) \times L^2(\Omega)$ :

$$(20) \quad \frac{\partial \lambda_1}{\partial t} + \Delta \lambda_1 + \gamma'(y) \lambda_1 = 0, \quad \lambda_1(T, \cdot) = 2(y(T, \cdot) - y_d), \quad \lambda_0 = \lambda_1(0, \cdot).$$

Of course, (20) is to be understood again in the weak sense.

To evaluate the Hamiltonian, let us observe that both  $A(\cdot, y)$  and  $\varphi(\cdot, y)$  are affine with respect to the geometry of  $\mathcal{F}_{(16)}^*$ , so that the Gâteaux differential of the extended mappings does not explicitly depend on particular relaxed controls and equals just  $A(\cdot, y)$  and  $\varphi(\cdot, y)$ , respectively. Then the abstract Hamiltonian  $f_{\lambda, y, v} \in \mathcal{F}_{(16)}$  of (7), after the transformation via  $\Psi_{(0, T)}^{-1}$  to  $\mathcal{H}_{\lambda, y, v_0} \equiv \mathcal{H}_{\lambda_1} \in L^1(0, T; C(B_\Omega^r))$ , looks like

$$(21) \quad \mathcal{H}_{\lambda_1}(t, v) = \int_{\Omega} \left( v(x) - \frac{1}{|\Omega|} \int_{\Omega} v(\zeta) d\zeta \right)^2 dx + \int_{\Omega} \lambda_1(t, x) \gamma(v(x)) dx + c(t),$$

where  $t \in (0, T)$  and  $v \in B_\Omega$ , and  $c \in L^1(0, T)$  is arbitrary. The Pontryagin maximum principle for the relaxed problem then looks like

$$\int_{B_\Omega} \mathcal{H}_{\lambda_1}(t, v) \nu_t(dv) = \sup_{v \in B_\Omega} \mathcal{H}_{\lambda_1}(t, v) \quad \text{for a.a. } t \in (0, T),$$

where  $\nu = \{\nu_t\}_{t \in (0, T)} \in L_w^\infty(0, T; rba(B_\Omega^r))$  is an optimal relaxed control.

Thus we have shown that our fairly abstract theory coincides with standard approaches in concrete cases.

## References

1. N.U. Ahmed, *Optimal control of a class of strongly nonlinear parabolic systems*, J. Math. Anal. Appl. **61** (1977), 188-207.
2. N.U. Ahmed, *Properties of relaxed trajectories for a class of nonlinear evolution equations on a Banach space*, SIAM J. Control Optim. **21** (1983), 953-967.
3. N.U. Ahmed and K.L. Teo, *Necessary conditions for optimality of Cauchy problems for parabolic partial differential systems*, SIAM J. Control **13** (1975), 981-993.
4. N. Basile and M. Mininni, *An extension of the maximum principle for a class of optimal control problems in infinite-dimensional spaces*, SIAM J. Control Optim. **28** (1990), 1113-1135.
5. A.G. Butkovskii, *Distributed Control Systems*, Elsevier, New York, 1969.
6. A.G. Chentsov, *Finitely additive measures and minimum problems*, (In Russian) Kibernetika No.3 (1988), 67-70.
7. A.G. Chentsov, *On the construction of solution to nonregular problems of optimal controls*, Problems Control Inf. Th. **20** (1991), 129-143.
8. H.O. Fattorini: *Relaxed controls in infinite dimensional control systems*, In: Proc. 5th Conf. on Control and Estimation of Distributed Parameter Systems, ISNM **100**, Birkhäuser Verlag, Basel, 1991.
9. H.O. Fattorini: *Relaxation theorems, differential inclusions and Filippov's theorem for relaxed controls in semilinear infinite dimensional systems*, J. Diff. Equations (to appear).
10. H.O. Fattorini: *Existence theory and the maximum principle for relaxed infinite dimensional optimal control problems*, SIAM J. Control Optim. (to appear).
11. H.O. Fattorini: *Relaxation in semilinear infinite dimensional control systems*, Markus Festschrift Volume, Lecture Notes in Pure and Appl. Math. **152**, 1993.
12. H.O. Fattorini and H. Frankowska, *Necessary conditions for infinite dimensional control problems*, (to appear).
13. H. Frankowska, *The maximum principle for differential inclusions with end point constraints*, SIAM J. Control **25** (1987).
14. R.V. Gamkrelidze: *Principles of Optimal Control Theory*, (in Russian), Tbilisi Univ. Press, Tbilisi, 1977. Engl. Transl. Plenum Press, New York, 1978.
15. W. Kampowsky, *Optimal control für Prozesse in Evolutionsgleichungen 1. Ordnung*, ZAMM **61** (1981), 401-407.
16. D.G. Luenberger: *Optimization by Vector Space Methods*, J. Wiley, New York, 1969.
17. N.S. Papageorgiou, *Optimal control of nonlinear evolution inclusions*, J. Optim. Th. Appl. **67** (1990), 321-354.
18. A.B. Pashaev and A.G. Chentsov, *Generalized control problem in a class of finitely-additive measures*, (in Russian) Kibernetika No.2 (1986), 110-112.
19. L.S. Pontryagin, V.G. Boltjanski, R.V. Gamkrelidze, E.F. Mischenko: *Mathematical Theory of Optimal Processes*, (in Russian), Moscow, 1961.
20. T. Roubíček, *Convex compactifications and special extensions of optimization problems*, Nonlinear Analysis, Th., Methods, Appl. **16**, 1117-1126 (1991).
21. T. Roubíček, *A general view to relaxation methods in control theory*, Optimization **23** (1992), 261-268.
22. T. Roubíček, *Minimization on convex compactifications and relaxation of nonconvex variational problems*, Advances in Math. Sciences and Appl. **1**, 1-18 (1992).
23. T. Roubíček, *Convex compactifications in optimal control theory*, In: Proc. 15th IFIP Conf. "Modelling and Optimization of Systems" (Ed. P. Kall), Lecture Notes in Control and Inf. Sci. **180** (1992), Springer, Berlin, pp. 433-439.
24. T. Roubíček, *Relaxation in Optimization Theory and Variational Calculus*, In preparation.
25. R. Sadigh-Esfandiari, J.M. Sloss, and J.C. Bruch Jr., *Maximum principle for optimal control of distributed parameter systems with singular mass matrix*, J. Optim. Th. Appl. **66** (1990),



211-226.

26. J.M. Sloss, J.C. Bruch Jr. and I.S. Sadek, *A maximum principle for nonconservative self-adjoint systems*, IMA J. Math. Control Inf. **6** (1986), 199-216.
27. T. Zolezzi, *Necessary conditions for optimal controls of elliptic or parabolic problems*, SIAM J. Control **10** (1972), 594-602.

Tomáš Roubíček  
Institute of Information Theory and Automation  
Academy of Sciences of Czech Republic  
Pod vodárenskou věží 4  
CZ-182 08 Praha 8,  
Czech Republic

# CONVERGENCE OF AN SQP-METHOD FOR A CLASS OF NONLINEAR PARABOLIC BOUNDARY CONTROL PROBLEMS

FREDI TRÖLTZSCH

Department of Mathematics  
Technical University of Chemnitz-Zwickau

**ABSTRACT.** We investigate local convergence of an SQP method for an optimal control problem governed by a parabolic equation with nonlinear boundary condition. Sufficient conditions for local quadratic convergence of the method are discussed.

1991 *Mathematics Subject Classification.* 49M05, 49M40, 49K24

*Key words and phrases.* Sequential quadratic programming, optimal control, parabolic equation, nonlinear boundary condition, control constraints.

## 1. Introduction

In this paper, we investigate the behaviour of a Sequential Quadratic Programming (SQP) method applied to the following very simplified model problem (P):

(P) Minimize

$$(1.1) \quad \frac{1}{2} \int_0^T \int_{\Gamma} \{ (w(t, \xi) - q(t, \xi))^2 + \lambda u(t, \xi)^2 \} dS_{\xi} dt$$

subject to

$$(1.2) \quad \begin{aligned} w_t(t, \xi) &= (\Delta_{\xi} w - w)(t, \xi) && \text{in } \Omega \\ w(0, \xi) &= 0 && \text{in } \Omega \\ \frac{\partial w}{\partial n}(t, \xi) &= b(w(t, \xi)) + u(t, \xi) && \text{on } \Gamma \end{aligned}$$

and to the constraints on the control

$$(1.3) \quad |u(t, \xi)| \leq 1,$$

$t \in [0, T]$ . The control  $u$  is looked upon in  $L_{\infty}((0, T) \times \Gamma)$ , while the state  $w$  is defined as mild solution of (1.2) (cf. section 2).

In this setting, a bounded domain  $\Omega \subset \mathbb{R}^n$  ( $n \geq 2$ ) with  $C^\infty$  boundary  $\Gamma$ , positive constants  $\lambda, T$ , and functions  $b \in C^2(\mathbb{R}), q \in L_\infty((0, T) \times \Gamma)$  are given. By  $n$  and  $dS$  the outward normal vector and the surface measure on  $\Gamma$ , respectively, are denoted.

Pointwise constraints of the type (1.3) are often imposed for a correct modelling of the underlying process. They reflect technical limitations to the possible choice of the control and cannot be realized by a smooth penalization. The term  $\lambda \|u\|^2$  in (1.1) is to enhance continuity of the optimal control (it may express the cost for the control, too).

It is known since several years that the SQP algorithm, applied to mathematical programming problems in finite-dimensional spaces, exhibits local quadratic convergence. The method can be easily extended to infinite-dimensional optimization problems such as optimal control problems. We refer, for instance, to the works by Alt [1], [2], Alt and Malanowski [3], Kelley and Wright [7], or Levitin and Polyak [10]. In the context of nonlinear parabolic control problems without control constraint we mention the numerical work by Kupfer and Sachs [8].

Recently, Alt, Sontag and Tröltzsch [4] proved the local quadratic convergence of the SQP method for the optimal control of a weakly singular Hammerstein integral equation with pointwise constraints on the control. In the author's paper [12], the proof of convergence was transferred to the one-dimensional heat equation with nonlinear boundary condition. The aim of this note is to extend the convergence result to a parabolic equation in a domain of dimension  $n$ .

We assume that  $b$  and its derivatives up to the order 2 are uniformly bounded and Lipschitz: There are constants  $c_B, c_l$ :

$$(1.4) \quad |b^{(i)}(w)| \leq c_B \quad |b^{(i)}(w_1) - b^{(i)}(w_2)| \leq c_l,$$

for all  $w, w_1, w_2 \in \mathbb{R}, i = 0, 1, 2$ . We may weaken the conditions (1.4) to local ones. However, this would lead to difficulties, as the solution  $w$  of (1.2) could blow up in finite time. For convenience we consider only the very special type of boundary condition in (1.2). This enables us to work out the principal behaviour of the SQP method and to avoid tedious technical estimates. The case of more general boundary conditions having the form  $\partial w / \partial n = b_1(w) + b_2(w)u$  is discussed in [4].

## 2. Integral equation method

Let us define  $A : L_2(\Omega) \supset D(A) \rightarrow L_2(\Omega)$  by  $D(A) = \{w \in W_2^2(\Omega) : \partial w / \partial n = 0 \text{ on } \Gamma\}$ ,  $Aw = -\Delta w + w$  for  $w \in D(A)$ .  $-A$  is known to generate an analytic semigroup  $\{S(t)\}, t \geq 0$ , of continuous linear operators in  $L_2(\Omega)$ . Moreover, we introduce the Neumann operator  $N : L_2(\Gamma) \rightarrow W_2^s(\Omega)$  ( $s < 1 + 1/2$ ) by  $N : g \mapsto w, \Delta w - w = 0, \partial w / \partial n = g$ . Next, we fix  $\sigma, p \in \mathbb{R}$  by  $p > n + 1$  and  $n/p < \sigma < 1 + 1/p$ . The last two inequalities have a non-void intersection for

$p > n - 1$ , while  $p > n + 1$  is needed to work with states being continuous w.r. to  $t$ . In the same way we may introduce operators  $A_r, S_r(t), N_r$  just by substituting above the order of integrability  $r$  for 2, where  $1 < r < \infty$ . Restricting  $AS(t)N$  to  $L_r(\Gamma)$  ( $r \geq 2$ ), we obtain  $A_r S_r(t) N_r$ . Therefore, we shall use in the paper the same symbol  $AS(t)N$  regarded in different spaces  $L_r$ . To continue our preparations we define a Nemytskii operator  $\mathcal{B}: C(\Gamma) \rightarrow C(\Gamma)$  by  $(\mathcal{B}x)(\xi) = b(x(\xi))$ .

A function  $w \in C([0, T], W_p^\sigma(\Omega))$  is said to be a *mild solution* of (1.2), if the Bochner integral equation

$$(2.1) \quad w(t) = \int_0^t AS(t-s)N(\mathcal{B}(\tau w(s)) + u(s))ds$$

holds on  $[0, T]$  ( $\tau$ : trace operator). The expression on the right hand side makes sense, as  $u \in L_\infty((0, T) \times \Gamma) \subset L_p((0, T) \times \Gamma) = L_p(0, T; L_p(\Gamma))$  and  $W_p^\sigma(\Omega) \subset C(\bar{\Omega})$  by  $n/p < \sigma$  (here we regard  $AS(t)N$  as operator from  $L_p(\Gamma)$  to  $W_p^\sigma(\Omega)$ ,  $t > 0$ ). Owing to the strong assumption (1.4), to each  $u \in L_\infty((0, T) \times \Gamma)$  a unique global solution  $w$  of (2.1) exists (cf. Tröltzsch [11]). Turning over to the trace  $x(t) = \tau w(t)$  in (2.1) we arrive at the integral equation

$$(2.2) \quad x(t) = \int_0^t \tau AS(t-s)N(\mathcal{B}(x(s)) + u(s))ds$$

for  $x \in C([0, T], C(\Gamma))$ .

The estimate (see Amann [4])

$$(2.3) \quad \|AS(t)N\|_{L_r(\Gamma) \rightarrow W_p^\sigma(\Omega)} \leq c t^{-\alpha},$$

where  $\alpha = 1 - (\sigma' - \sigma)/2$  and  $0 < \sigma < \sigma' < 1 + 1/r$ , turns out to be essential for investigating properties of the integral operator  $K$ .

$$(2.4) \quad (Kz)(t) = \int_0^t \tau AS(t-s)Nz(s)ds.$$

Let us briefly discuss (2.3) for  $r := p$ : Taking  $\sigma = n/p + \varepsilon, \sigma' = 1 + 1/p - \varepsilon$  ( $\varepsilon > 0$ ) we find that (2.3) holds for all  $\alpha > 0.5 + (n-1)/2p$ .  $K$  maps continuously  $L_p(0, T; L_p(\Gamma))$  into  $C([0, T], C(\Gamma))$  provided that  $p > 1/(1-\alpha)$ . This holds together with the last inequality for  $\alpha$ , if  $p > n+1$ . For  $p \downarrow n+1$  we may take  $\alpha \downarrow n/(n+1)$ .

For convenience we regard  $K$  between different spaces:  $K$  may be viewed as operator in  $L_r(0, T; L_r(\Gamma))$  for all  $1 < r < \infty$ . Let its adjoint  $K^*$  be defined for  $r = 2$ .

It is known that

$$(K^*z)(t) = \int_t^T \tau A S(s-t) N z(s) ds,$$

hence  $K^*$  has the same transformation properties as  $K$ . By means of these prerequisites we are able to write (P) as

$$f(x, u) = \frac{1}{2} \int_0^T \{ \|x(t) - q(t)\|_{L_2(\Gamma)}^2 + \lambda \|u(t)\|_{L_2(\Gamma)}^2 \} dt = \min! \quad (P)$$

subject to

$$(2.5) \quad x = K(B(x) + u), \quad u \in U^{ad}.$$

Here we have introduced  $B$  in  $C([0, T], C(\Gamma))$  by  $(Bx)(t, \cdot) = \mathcal{B}(x(t, \cdot))$  and  $U^{ad} := \{u \in L_\infty((0, T) \times \Gamma) : |u(t, \xi)| \leq 1\}$ . The equation in (2.5) is well defined, as  $K$  maps  $L_p(0, T; L_p(\Gamma))$  into  $C([0, T], C(\Gamma))$ .

In the paper, the following notation is used: We write  $L_r = L_r(0, T; L_r(\Gamma))$ ,  $1 \leq r < \infty$ ,  $L_\infty = L_\infty((0, T) \times \Gamma)$ ,  $C = C([0, T], C(\Gamma)) = C([0, T] \times \Gamma)$  and endow the spaces with their natural norms  $\|\cdot\|_r$  and  $\|\cdot\|_\infty$ , respectively. The natural norm of  $L_\alpha(0, T; L_\beta(\Gamma))$ ,  $1 < \alpha, \beta < \infty$ , is denoted by  $\|\cdot\|_{L_{\alpha, \beta}}$ . For  $\alpha = \beta = \infty$  we set  $\|\cdot\|_{L_{\infty, \infty}} := \|\cdot\|_\infty$ . In product spaces of this type, the norm is defined as the sum of the corresponding norms. In  $C \times L_r$ ,  $\|(x, u)\|_{\infty, r} = \|x\|_\infty + \|u\|_r$ ,  $1 \leq r \leq \infty$ , and  $\|(x, u)\|_r := \|(x, u)\|_{r, r}$ , in  $C \times L_r \times L_r$ :  $\|(x, u, y)\|_{\infty, r} = \|x\|_\infty + \|(u, y)\|_r$ ,  $\|(x, u, y)\|_\infty := \|(x, u, y)\|_{\infty, \infty}$ . An "inner product" is defined formally by

$$\langle x, y \rangle = \int_0^T \int_\Gamma x(t, \xi) y(t, \xi) dS_\xi dt$$

just denoting integration of  $xy$  over  $[0, T] \times \Gamma$ .

### 3. Known optimality conditions

The functional  $f : C \times L_p \rightarrow \mathbb{R}$  and the mapping  $(x, u) \mapsto B(x) + u$  from  $C \times L_p$  to  $L_p$  are twice continuously Fréchet differentiable. This enables us to apply lateron second order methods to (P). Owing to the convexity of  $f$  and the linear appearance of  $u$  in (1.2), standard methods show the existence of at least one optimal control  $u_0$  for (P). Let  $x_0$  be the corresponding state. For  $y \in L_\infty$  the Lagrange function

$$\mathcal{L}(x, u, y) = f(x, u) - \langle y, x - K(B(x) + u) \rangle$$

is defined. From  $\mathcal{L}_x = 0$  we obtain formally the equation  $y_0 = f_x + B'(x_0)^* K^* y_0$  for the Lagrange multiplier  $y_0$ . A careful discussion (taking derivatives in  $C \times L_\infty$

and regarding the derivatives as linear operators in  $L_p$ ) justifies this (cf. Tröltzsch [11]):

$$(3.1) \quad y_0(t) = x_0(t) - q(t) + B'(x_0(t)) \int_t^T \tau AS(s-t)N y_0(s) ds.$$

Here, we took advantage of  $(B'(x_0(t, \cdot))h(t, \cdot))(\xi) = b'(x_0(t, \xi)) \cdot h(t, \xi)$ , hence  $B'(x_0)$  is formally self-adjoint. The variational inequality  $\langle \mathcal{L}_u, u - u_0 \rangle \geq 0 \forall u \in U^{nd}$  yields  $\langle \lambda u_0 + K^* y_0, u - u_0 \rangle \geq 0$ . After a standard pointwise discussion we arrive at

$$(3.2) \quad u_0(t, \xi) = P_{[-1,1]} \{-\lambda^{-1}(K^* y_0)(t, \xi)\},$$

where  $P_{[-1,1]} : \mathbb{R} \rightarrow [-1,1]$  denotes projection onto  $[-1,1]$ . We assume that in addition to the first order necessary conditions (3.1), (3.2) the following *second order sufficient optimality condition* is satisfied: There is a  $\delta > 0$  such that

$$(3.3) \quad \mathcal{L}_{vv}(x_0, u_0, y_0)[v - v_0, v - v_0] \geq \delta \|v - v_0\|_2^2 \quad (\text{SSC})$$

for all  $v = (x, u)$  satisfying the linearized equation

$$(3.4) \quad x - x_0 = K(B'(x_0)(x - x_0) + u - u_0).$$

In (3.3),  $\mathcal{L}_{vv}$  denotes the second order  $F$ -derivative of  $\mathcal{L}$  w.r. to  $v = (x, u)$  in  $C \times L_p$  at  $v_0 := (x_0, u_0)$ . The sufficiency of (SSC) was discussed in [6]. We finish this section by the following very useful Lemma:

**Lemma 3.1.** *Let  $1 \leq r, \rho \leq \infty$ ,  $D$  be the linear continuous operator in  $L_{r,\rho}$  defined by  $(Dx)(t, \xi) = \beta(t, \xi)x(t, \xi)$ , where  $\beta \in L_\infty$ . Then  $(I - DK)^{-1}$  and  $(I - KD)^{-1}$  exist as linear continuous operators in  $L_{r,\rho}$  and their norm is bounded by a constant  $c_{r,\rho}$  which depends only  $\|\beta\|_\infty$ ,  $r$ , and  $\rho$ .*

**Proof.** The invertibility follows from a standard application of the Banach fixed point theorem. Let  $z \in L_{r,\rho}$  be given. We estimate the solution of the equation  $x = BKx + z$ . Then

$$\begin{aligned} \|x(t)\|_{L_\rho(\Gamma)} &\leq \int_0^t \|\beta\|_\infty \|\tau AS(t-s)N\|_{L_\rho(\Gamma) \rightarrow L_\rho(\Gamma)} \|x(s)\|_{L_\rho(\Gamma)} ds + \|z(t)\|_{L_\rho(\Gamma)} \\ &\leq c \int_0^t (t-s)^{-\alpha} \|x(s)\|_{L_\rho(\Gamma)} ds + \|z(t)\|_{L_\rho(\Gamma)}. \end{aligned}$$

Here,  $c > 0$  depends only on  $\rho$ ,  $\|\beta\|_\infty$ , and  $\alpha \in (0, 1)$ .  $\|x(t)\|$  is majorized by the real function  $\varphi(t)$  solving the corresponding weakly singular integral equation. Therefore

$$\|x\|_{L_{r,\rho}} = \left( \int_0^T \|x(t)\|_{L_\rho(\Gamma)}^r dt \right)^{1/r} \leq \|\varphi\|_{L_r(0,T)} \leq c \left( \int_0^T \|z(t)\|_{L_\rho(\Gamma)}^r dt \right)^{1/r} = c \|z\|_{L_{r,\rho}}.$$

In this way the Lemma is shown for  $I - DK$ . The arguing for  $I - KD$  is identical.  $\square$

#### 4. The SQP method, Hölder estimates

Initiating from a starting point  $(x_1, u_1, y_1)$  in  $C \times L_\infty \times L_\infty$  the (full) SQP method generates sequences  $\{x_n\}, \{u_n\}, \{y_n\}$  by solving certain quadratic programs. Adopting the notation by Alt [1], one step of the method can be described as follows: Let  $w = (x_w, u_w, y_w)$  be the result of the last iteration. As before, we write  $v_w = (x_w, u_w)$ ,  $v = (x, u)$ . The next iterate  $\bar{v}_w = (\bar{x}_w, \bar{u}_w)$  is obtained as the solution of

$$(4.1) \quad F(v; w) = f'(v_w)(v - v_w) + \frac{1}{2} \mathcal{L}_{vv}(v_w, y_w)[v - v_w, v - v_w] = \min! \quad (QP)_w$$

$$(4.2) \quad x = K(B'(x_w)(x - x_w) + B(x_w) + u), \quad u \in U^{ad}.$$

$\bar{y}_w$  is the corresponding Lagrange multiplier.

*Remark.* Define  $g(v) = g(x, u) = x - K(B(x) + u)$ . Then the state equation (2.5) reads  $g(v) = 0$ . (4.2) is its linearization  $g(v_w) + g'(v_w)(v - v_w) = 0$ , which simplifies by linearity w.r. to  $u$ .

The Lagrange function  $\tilde{\mathcal{L}}$  to  $(QP)_w$  is

$$\tilde{\mathcal{L}}(v, y) = F(v; w) - \langle y, x - K(B'(x_w)(x - x_w) + B(x_w) + u) \rangle.$$

$\tilde{\mathcal{L}}_x = 0$  leads to the (adjoint) equation

$$(4.3) \quad y = B'(x_w)K^*y + f_x(v_w) + f_{x,x}(v_w)[\bar{x}_w - x_w, \cdot] + B'(x_w)[\bar{x}_w - x_w, \cdot]K^*y_w$$

for  $y = \bar{y}_w$ . After a simple calculation we find

$$(4.4) \quad \begin{aligned} \bar{y}_w(t) = & B'(x_w(t)) \int_t^T \tau AS(s-t) N \bar{y}_w(s) ds + \bar{x}_w(t) - q(t) \\ & + B''(x_w(t)) [\bar{x}_w(t) - x_w(t), \cdot] \int_t^T \tau AS(s-t) N y_w(s) ds. \end{aligned}$$

The variational inequality determining  $\bar{u}_w, \langle \tilde{\mathcal{L}}_w, u - \bar{u}_w \rangle \geq 0 \quad \forall u \in U^{ad}$ , gives

$$(4.5) \quad \bar{u}_w(t, \xi) = P_{[-1, 1]} \{-\lambda^{-1}(K^* \bar{y}_w)(t, \xi)\}.$$

A straightforward calculation by means of (SSC) yields

$$(4.6) \quad F(v; w_0) \geq \delta \|v - v_0\|_2^2 = F(v_0; w_0) + \delta \|v - v_0\|_2^2$$

for all  $v$  being admissible for  $(QP)_{w_0}$  (where  $w_0 = (v_0, y_0) = (x_0, u_0, y_0)$ ).

**Lemma 4.1.** *For all  $w$  in a sufficiently small  $C \times L_p \times L_p$  neighbourhood  $N_1(w_0)$ ,  $(QP)_w$  admits a unique solution  $(\bar{x}_w, \bar{u}_w) = \bar{v}_w$  with associated Lagrange multiplier  $\bar{y}_w$ . There is a constant  $c_H$  not depending on  $w$ , such that*

$$(4.7) \quad \|\bar{v}_w - v_0\|_2 \leq c_H \|w - w_0\|_2^{1/2}$$

$$(4.8) \quad \|\bar{y}_w\|_\infty \leq c_H$$

for all  $w \in N_1(w_0)$ .

**Proof.** We shall only briefly sketch the proof, which is along the lines of Alt [1] or Alt, Sontag and Tröltzsch [3]. A first step initiates from the simple observation  $F(\bar{v}_w; w) \leq F(\tilde{v}_w; w)$ , where  $\tilde{v}_w = (\tilde{x}_w, u_w)$  and  $\tilde{x}_w$  is the state obtained from (4.2) for  $u = u_w$ . By means of Lemma 3.1 the *upper estimate*

$$(4.9) \quad F(\bar{v}_w; w) \leq c \|w - w_0\|_2$$

can be derived. A *lower estimate*

$$(4.10) \quad F(\bar{v}_w; w) \geq \frac{\delta}{2} \|\tilde{v}_w - v_0\|_2^2 - c \|w - w_0\|_2$$

follows from re-writing terms such as  $f'(v_w)$  or  $\mathcal{L}_{cv}(v_w, y_w)$  in terms of  $f'(v_0)$ ,  $\mathcal{L}_{cv}(v_0, y_0)$  etc., estimating the correction parts and exploiting (SSC) and Lemma 3.1. We omit the details of the lengthy computations. (4.9) and (4.10) yield (4.7). As regards (4.8), we first observe that  $\bar{u}_w$  is uniformly bounded, hence  $\|\bar{x}\|_\infty$  is bounded, too (this is a consequence of (4.2) for  $u = \bar{u}_w$ :  $\|x_w\|_\infty$  is bounded, as  $w \in N_1(w_0)$ ; apply Lemma 3.1 to  $x = \bar{x}_w$  in (4.2)). Now the uniform boundedness of  $\bar{y}_w$  is an immediate conclusion of (4.4) and Lemma 3.1.  $\square$

**Corollary 4.2.** *The estimate (4.7) holds true in the form*

$$(4.11) \quad \|\bar{v}_w - v_0\|_{\infty, p} \leq c'_H \|w - w_0\|_{\infty, p}^{1/p}$$

for all  $w \in N'_1(w_0) \subset N_1(w_0)$ .



**Proof.** (4.7) means in particular  $\|\bar{u}_w - u_0\|_2 \leq c_H \|w - w_0\|_2^{1/2}$ . Exploiting  $|\bar{u}_w - u_0| \leq 2$  it is easy to show that

$$(4.12) \quad \|\bar{u}_w - u_0\|_p \leq c \|w - w_0\|_2^{1/p} \leq c' \|w - w_0\|_{\infty,p}^{1/p}.$$

Subtracting the equations for  $\bar{x}_2$  and  $x_0$  we get

$$(\bar{x}_w - x_0) - KB'(x_w)(\bar{x}_w - x_0) = K(B(x_w) - B(x_0) + B'(x_w)(x_0 - x_w) + \bar{u}_w - u_0).$$

Therefore, the  $L_\infty$  norm of the left hand side is less or equal  $\|K\|_{L_p \rightarrow C}(c_1 \|x_w - x_0\|_\infty + c_2 \|\bar{u}_w - u_0\|_p) \leq c \|w - w_0\|_{\infty,p}^{1/p}$  by (4.12) (provided that  $\|w - w_0\|_{\infty,p} \leq 1$ ). Now Lemma 3.1 applies to the left hand side,

$$\|\bar{x}_w - x_0\|_\infty \leq c'' \|w - w_0\|_{\infty,p}^{1/p},$$

implying (4.11).  $\square$

In the same way, subtraction of the equations for  $y_0, y_w$  yields

**Corollary 4.3.** *There is a constant  $c_H'' > 0$  such that*

$$(4.13) \quad \|\bar{y}_w - y_0\|_\infty \leq c_H'' \|w - w_0\|_{\infty,p}^{1/p}$$

for all  $w \in N_1''(w_0) \subset N_1(w_0)$ .

We omit the proof. In what follows let  $N_1(w_0)$  denote the intersection  $N_1(w_0) \cap N_1''(w_0) \cap N_1'''(w_0)$ .

## 5. Right hand side perturbations, Lipschitz estimate

Following Alt [1], [2], we consider now the close relationship between the stability of  $(QP)_w$  and certain perturbations of  $(QP)_{w_0}$ . We discuss the perturbed problem

$$f'(v_0)(v - v_0) + \frac{1}{2} \mathcal{L}_{vv}(v_0, y_0)[v - v_0, v - v_0] = \langle d, v - v_0 \rangle = \min! \quad (QS)_\pi$$

$$(5.1) \quad x = v + K(B'(x_0)(x - x_0) + B(x_0) + u), \quad u \in U^{ad}$$

belonging to the perturbation  $\pi = (d, v) = (d_x, d_u, v) \in L_\infty \times L_\infty \times C$ . For  $\pi = 0$  this problem has the unique solution  $v_0 = (x_0, u_0)$ . In  $(QS)_\pi$  we regard  $x, u$  in  $L_2$ , although the constraint  $u \in U^{ad}$  automatically generates only  $L_\infty$ -solutions.

$(QS)_\pi$  is a linear-quadratic parabolic control problem, where the theory is already widely investigated. Owing to (SSC), the following result is therefore standard: There is a neighbourhood  $N_2(0)$ , and a positive constant  $c_h > 0$  such that for

all  $c \in N_2(0)$  and all  $d \in L_\infty \times L_\infty$  problem  $(QS)_\pi$  admits a unique solution  $v_\pi = (x_\pi, u_\pi)$  and

$$(5.2) \quad \|v_\pi - v_0\|_2 \leq c_h \|\pi\|_2^{1/2}$$

for all  $\pi = (d, c)$  such that  $c \in N_2(0)$ . We are able to improve this estimate in Theorem 5.3.

**Lemma 5.1.** *Let  $y_\pi$  be the Lagrange multiplier belonging to  $v_\pi$  and  $2 \leq \alpha, \beta \leq \infty$ . Then there is a constant  $c_{\alpha\beta} > 0$  such that*

$$(5.3) \quad \|y_\pi - y_0\|_{L_{\alpha,\beta}} \leq c_{\alpha\beta} (\|x_\pi - x_0\|_{L_{\alpha,\beta}} + \|\pi\|_{L_{\alpha,\beta}})$$

for all  $\pi \in L_\infty \times L_\infty \times C$ .

**Proof.** The adjoint equations defining  $y_0, y_\pi$  are (3.1) and

$$(5.4) \quad y_\pi = (x_\pi - q) - d_x + B'(x_0)K^*y_\pi + B''(x_0)[x_\pi - x_0, \cdot]K^*y_0.$$

Subtraction of these equations yields after some estimations

$$\|(y_0 - y_\pi) - B'(x_0)K^*(y_0 - y_\pi)\|_{L_{\alpha,\beta}} = \|d_x + x_\pi - x_0 + B''(x_0)[x_\pi - x_0, \cdot]K^*y_0\|_{L_{\alpha,\beta}}.$$

Applying Lemma 3.1

$$\|y_0 - y_\pi\|_{L_{\alpha,\beta}} \leq c_1 \|d_x\|_{L_{\alpha,\beta}} + c_2 \|x_\pi - x_0\|_{L_{\alpha,\beta}}$$

is obtained. This implies (5.3).  $\square$

One of the decisive steps for showing quadratic convergence is the following Lipschitz estimate improving (5.2):

**Theorem 5.2.** *There is a constant  $c_L > 0$  such that*

$$(5.5) \quad \|v_\pi - v_0\|_2 \leq c_L \|\pi\|_2$$

for all  $\pi \in L_\infty \times L_\infty \times C$ .

**Proof.** We outline the main steps of the proof. The first order condition for  $v_\pi$  as a solution of  $(QS)_\pi$  is

$$0 \leq \langle \tilde{\mathcal{L}}_v(v_\pi, y_\pi), v - v_\pi \rangle \quad \forall v \in L_2 \times U^{ad},$$

i.e.

$$0 \leq f'(v_0)(v - v_\pi) + \mathcal{L}_{vv}(v_0, y_0)[v_\pi - v_0, v - v_\pi] - \langle d, v - v_\pi \rangle - \langle y_\pi, (x - x_\pi) - K(B'(x_0)(x - x_\pi) + u - u_\pi) \rangle$$

for all  $x \in L_2$ ,  $u \in L^{ad}$ . Now we insert  $x = x_0$ ,  $v = (x_0, u_0)$ ,  $u = u_0$  and find after exploiting the first order necessary optimality conditions for  $v_0$  as a solution for (P)

$$\begin{aligned} \mathcal{L}_{vv}(v_0, y_0)[v_\pi - v_0, v_\pi - v_0] &\leq -\langle c, y_0 - y_\pi \rangle - \langle d, v_0 - v_\pi \rangle \\ (5.6) \quad &\leq \|y_0 - y_\pi\|_2 \|\pi\|_2 + \|d\|_2 \|v_0 - v_\pi\|_2 \\ &\leq c \|\pi\|_2^2 + c \|\pi\|_2 \|v_0 - v_\pi\|_2 \end{aligned}$$

by Lemma 5.1. The difference  $\xi = v_\pi - v_0 = (x_\pi - x_0, u_\pi - u_0) = (\xi_x, \xi_u)$  solves  $\xi_x - K(B'(x_0)\xi_x + \xi_u) = c$ , hence  $\xi$  does not satisfy the linearized equation (3.4), where (SSC) applies. Define  $\hat{\xi} = (\hat{\xi}_x, \xi_u)$ , where  $\hat{\xi}_x$  is the solution of  $\hat{\xi}_x - K(B'(x_0)\hat{\xi}_x + \xi_u) = 0$ . Then  $\xi_x = \hat{\xi}_x + \Delta$ , and  $\|\Delta\|_2 \leq c \|\pi\|_2$  (apply Lemma 3.1). (SSC) is valid for  $\hat{\xi}_x$ , hence simple estimations yield

$$\begin{aligned} \mathcal{L}_{vv}(v_0, y_0)[\xi, \xi] &\geq \delta \|\hat{\xi}\|_2^2 - 2c \|\hat{\xi}\|_2 \|\Delta\|_2 - c \|\Delta\|_2^2 \\ (5.7) \quad &\geq \delta \|\xi\|_2^2 - c(\|\xi\|_2 \|\Delta\|_2 + \|\Delta\|_2^2). \end{aligned}$$

Taking into account (5.6) and  $\|\Delta\|_2 \leq c \|\pi\|_2$  we easily find

$$\|\xi\|_2^2 \leq c(\|\xi\|_2 \|\pi\|_2 + \|\pi\|_2^2) \leq c \|\xi\|_2 \|\pi\|_2,$$

if  $\|\pi\|_2 \leq \|\xi\|_2$ . This implies (5.5), if  $\|\xi\|_2 \geq \|\pi\|_2$ . Thus (5.5) holds for  $c_L := \max(1, c)$ .  $\square$

The estimate (5.5) in the  $L_2$ -norm is not sufficient for our purposes. However, we are able to show

**Theorem 5.3.** *There is a constant  $c'_L > 0$  such that*

$$(5.8) \quad \|v_\pi - v_0\|_\infty \leq c'_L \|\pi\|_{\infty, p}$$

for all  $\pi \in L_\infty \times L_\infty \times C$ .

**Proof.** We start with the equation for  $x_\pi - x_0$ ,

$$(5.9) \quad x_\pi - x_0 - KB'(x_0)(x_\pi - x_0) = c + K(u_\pi - u_0).$$

We have  $K(u_\pi - u_0) = \tau w$ , where  $w$  solves the PDE (1.2) with boundary condition  $\partial w / \partial n = u_\pi - u_0$ . By  $L_2$ -regularity,

$$\|w\|_{L_2(0, T; H^{3/2-\varepsilon}(\Omega))} \leq c \|u_\pi - u_0\|_2,$$

( $\varepsilon > 0$  fixed sufficiently small), hence

$$(5.10) \quad \|K(u_\pi - u_0)\|_{L_2(0, T; H^{1-\varepsilon}(\Gamma))} \leq c \|u_\pi - u_0\|_2.$$

Sobolev embedding theorems yield

$$H^{1-\varepsilon}(\Gamma) \subset L_{\frac{2(n-1)}{n-1-2(1-\varepsilon)}}(\Gamma) = L_{\frac{n-1}{(n-1)/2-(1-\varepsilon)}}(\Gamma) =: L_{p_1}(\Gamma).$$

Denote the left hand side of (5.9) by  $E$ . Thus

$$\begin{aligned}\|E\|_{L_2, p_1} &\leq \|c\|_{L_2, p_1} + c\|u_\pi - u_0\|_2 \leq \|c\|_{L_2, p_1} + c\|\pi\|_2 \\ &\leq c\|\pi\|_{L_2, p_1} \quad (\text{Theorem 5.2}).\end{aligned}$$

By Lemma 3.1,

$$(5.11) \quad \|x_\pi - x_0\|_{L_2, p_1} \leq c\|E\|_{L_2, p_1} \leq c\|\pi\|_{L_2, p_1}.$$

Invoking the first order necessary conditions for  $u_\pi$ ,  $u_\pi = P_{[-1,1]}\{-\lambda^{-1}(K^*y_\pi - d_u)\}$ , a simple estimation yields

$$(5.12) \quad \begin{aligned}\|u_\pi - u_0\|_{L_2, p_1} &\leq \lambda^{-1}\|K^*\|\|y_\pi - y_0\|_{L_2, p_1} + \lambda^{-1}\|d_u\|_{L_2, p_1} \\ &\leq c\|x_\pi - x_0\|_{L_2, p_1} + c\|\pi\|_{L_2, p_1} \leq c\|\pi\|_{L_2, p_1}\end{aligned}$$

by Lemma 5.1 and (5.11). In this way, we have already extended (5.5) to the  $L_2(0, T; L_{p_1}(\Gamma))$ -norm performing one step of a bootstrapping argument. Now we continue estimating (5.9) by means of the  $L_{p_1}$ -regularity of parabolic equations. The solution  $w$  can be estimated in the  $L_2(0, T; W_{p_1}^{1+1/p_1-\varepsilon}(\Omega))$ -norm, hence we have for its trace  $\tau w = K(u_\pi - u_0)$

$$\|K(u_\pi - u_0)\|_{L_2(0, T; W_{p_1}^{1-\varepsilon}(\Gamma))} \leq c\|u_\pi - u_0\|_{L_2, p_1}.$$

Embedding  $W_{p_1}^{1-\varepsilon}(\Gamma) \subset L_{\frac{n-1}{(n-1)/p_1-(1-\varepsilon)}}(\Gamma) = L_{\frac{n-1}{(n-1)/2-k(1-\varepsilon)}}(\Gamma) = L_{p_2}(\Gamma)$

$$\|x_\pi - x_0\|_{L_2, p_2} \leq c\|\pi\|_{L_2, p_2}$$

is obtained as above. Proceeding in the same way we arrive after at most  $\lfloor (n-1)/2 \rfloor + 1$  steps at the case  $(n-1)/2 - k(1-\varepsilon) < 0$  while  $(n-1)/2 - (k-1)(1-\varepsilon) > 0$  (provided  $\varepsilon > 0$  is sufficiently small). Here we end up with the possibility of an estimate in the norm of  $L_2(0, T; C(\Gamma))$ . However, we use only

$$\begin{aligned}\|K(u_\pi - u_0)\|_{L_2, p} &\leq c\|K(u_\pi - u_0)\|_{L_2(0, T; W_{p_{k-1}}^{1-\varepsilon}(\Gamma))} \\ &\leq c\|\pi\|_{L_2, p_{k-1}} \leq c\|\pi\|_{L_2, p},\end{aligned}$$

provided that  $p_{k-1} \leq p$ . If  $p_{k-1} > p$ , then we use the argument

$$\|K(u_\pi - u_0)\|_{L_2, p} \leq c\|K(u_\pi - u_0)\|_{L_2, p_{k-1}} \leq c\|\pi\|_{L_2, p_{k-2}} \leq c\|\pi\|_{L_2, p}$$

(note that  $p_{k-2} < p$  must hold, as  $L_{2, p_{k-2}}$  is still not transformed into  $C$ ).

Thus finally (invoking (5.9) and the optimality conditions for  $u_\pi, u_0$ )

$$(5.13) \quad \|v_\pi - v_0\|_{L_2, p} \leq c\|\pi\|_{L_2, p}$$

can be derived.

It remains to lift the regularity with respect to the time  $t$ . From Krasnosel'skii a.o. [9] it is known that a weakly singular integral operator with weak singularity  $\alpha \in (0, 1)$  maps continuously  $L_p(0, T)$  into  $L_{p'}(0, T)$ , if  $1/p' > 1/p + \alpha - 1$ . Put

$\delta = 1 - \alpha > 0$  and take  $\lambda \in (0, 1)$ . Then  $K$  transforms  $L_2(0, T; L_p(\Gamma))$  into  $L_{\beta_1}(0, T; L_p(\Gamma))$ , where  $1/\beta_1 = 1/2 - \lambda\delta$ . Arguing as in the first part of the proof,

$$(5.14) \quad \|u_0 - u_\pi\|_{L_{\beta_1,p}} \leq c\|\pi\|_{L_{\beta_1,p}}$$

is obtained.  $K$  transforms  $L_{\beta_1}(0, T; L_p(\Gamma))$  into  $L_{\beta_2}(0, T; L_p(\Gamma))$ , provided that  $1/\beta_2 = 1/\beta_1 - \lambda\delta = 1/2 - 2\lambda\delta$ . Therefore, the estimate (5.14) can be derived in the norm of  $L_{\beta_2}(0, T; L_p(\Gamma))$ . Proceeding in this way, (5.14) is seen to hold in the  $\|\cdot\|_p$ -norm after finitely many steps. (5.8) follows easily, as  $K$  transforms  $L_p$  into  $C$ .  $\square$

The next two results are standard. We refer to the proofs given by Alt in [1], which simplify considerably for our model problem (P).

**Lemma 5.4.** *There is a  $C \times L_p \times L_p$ -neighbourhood  $N_3(w_0)$ , such that for all  $w \in N_3(w_0)$  the following equivalence holds true: If  $w \in N_3(w_0)$ , then the solution  $\bar{v}_w = (\bar{x}_w, \bar{u}_w)$  is also the unique solution of  $(QS)_\pi$  for the following choice of  $\pi = (d, c) = (d_x, d_u, c) : d_u = 0$ .*

$$(5.15) \quad \begin{aligned} d_x = & B''(x_0)[\bar{x}_w - x_0, \cdot]K^*y_0 - B''(x_w)[\bar{x}_w - x_w, \cdot]K^*y_w \\ & - (B'(x_w) - B'(x_0))K^*\bar{y}_w \end{aligned}$$

$$(5.16) \quad c = K(B'(x_w)(\bar{x}_w - x_w) - B'(x_0)(\bar{x}_w - x_0) + B(x_w) - B(x_0)).$$

**Proof.** We know that  $(QS)_\pi$  is a convex problem with a solution determined uniquely by the conditions (5.1), (5.4), and

$$(5.17) \quad u_\pi = P_{[-1,1]} \{-\lambda^{-1}(K^*y_\pi - d_u)\}.$$

Thus it suffices to show that the triplet  $(\bar{x}_w, \bar{u}_w, \bar{y}_w)$  fulfils these relations for an appropriate  $\pi$  and  $y_\pi := \bar{y}_w$ . As regards  $\bar{x}_w$ , it is a solution of

$$\bar{x}_w = K(B'(x_w)(\bar{x}_w - x_w) + B(x_w) + \bar{u}_w).$$

In order to comply with (5.1), it must hold

$$\bar{x}_w = K(B'(x_0)(\bar{x}_w - x_0) + B(x_0) + \bar{u}_w) + c.$$

Subtracting the last equations we end up with (5.16). The adjoint state  $\bar{y}_w$  is defined by (4.4). Comparing this with (5.4),

$$\bar{y}_w = \bar{x}_w - q - d_x + B'(x_0)K^*\bar{y}_w + B''(x_0)[\bar{x}_w - x_0, \cdot]K^*y_0,$$

we easily arrive at formula (5.15). Obviously,  $\bar{u}_w$  satisfies (4.5) together with (5.17) iff  $d_u = 0$ .  $\square$

**Lemma 5.5.** *Define  $d$  and  $c$  according to (5.15), (5.16). Then for all  $w \in N_A(w_0)$*

$$(5.18) \quad \|c\|_{\infty} \leq c_T(\|x_0 - x_w\|_{\infty}^2 + \|x_w - x_0\|_{\infty} \|\bar{x}_w - x_0\|_{\infty})$$

$$(5.19) \quad \|d\|_{\infty} \leq c_T(\|\bar{y}_w\|_p \|x_0 - x_w\|_{\infty}^2 + \|\bar{x}_w - x_0\|_{\infty} (\|x_w - x_0\|_{\infty} + \|y_w - y_0\|_p) + \|x_w - x_0\|_{\infty} (\|y_w - y_0\|_p + \|\bar{y}_w - y_0\|_p))$$

with a certain constant  $c_T$  not depending on  $w$ .

The proof follows completely analogous to [4] from re arranging and estimating (5.15) - (5.16).

## 6. Quadratic convergence of the SQP-method

**Theorem 6.1.** *There is a  $C \times L_p \times L_p$  neighbourhood  $N_5(w_0)$ , and a positive constant  $\nu$  such that for all  $w \in N_5(w_0)$  the solution  $\bar{v}_w$  of  $(QP)_w$  and the corresponding Lagrange multiplier  $\bar{y}_w$  satisfy*

$$(6.1) \quad \|(\bar{v}_w, \bar{y}_w) - (v_0, y_0)\|_{\infty, p} \leq \nu \|w - w_0\|_{\infty, p}^2.$$

**Proof.** We take at first  $N(w_0) \subset N_1(w_0) \cap N_q(w_0)$  such that the radius of  $N(w_0)$  is less than 1. According to Corollary 4.2.,  $\|\bar{x}_w - x_0\|_{\infty}$  and  $\|\bar{y}_w\|_p$  remain bounded by a constant  $c > 0$  for all  $w \in N(w_0)$ . From (5.15) - (5.16)

$$(6.2) \quad \begin{aligned} \max(\|c\|_{\infty}, \|d\|_{\infty}) &\leq c(\|v_0 - v_w\|_{\infty}^2 + \|w_0 - w_w\|_{\infty, p}) \\ &\leq c\|w - w_0\|_{\infty, p} \end{aligned}$$

as the diameter of  $N(w_0)$  is less than 1.

Thus on  $N(w_0)$ ,

$$(6.3) \quad \|\pi\|_{\infty} \leq c\|w - w_0\|_{\infty, p}.$$

On the other hand, Lemma 5.4. and Theorem 5.3. yield now  $y_{\pi} = y_{\bar{w}}$  and

$$(6.4) \quad \|\bar{x}_w - v_0\|_{\infty, p} \leq c\|\pi\|_{\infty, p} \leq c\|w - w_0\|_{\infty, p}.$$

Analogously we find

$$(6.5) \quad \|\bar{y}_w - y_0\|_p \leq c\|w - w_0\|_{\infty, p}$$

by Lemma 5.1. and (6.3), (6.4). Inserting (6.4) - (6.5) in (5.18) - (5.19) we obtain

$$(6.6) \quad \|\pi\|_{\infty, p} \leq c\|\pi\|_{\infty} = \|c\|_{\infty} + \|d\|_{\infty} \leq c\|w - w_0\|_{\infty, p}^2,$$

implying together with (6.4), (5.3) the relation (6.1).  $\square$

Now we reformulate the *SQP* method and state the result on its local convergence. The *SQP*-method runs as follows.

**(SQP):** Choose a starting point  $w_1 = (v_1, y_1)$ . Having  $w_k = (v_k, y_k)$  compute  $w_{k+1} = (v_{k+1}, y_{k+1})$  to be the solution and the associated Lagrange multiplier of the quadratic optimization problem  $(QP)_{w_k}$ .

Using Theorem 6.1, it follows now by standard proof techniques that this method converges quadratically to  $w_0 = (x_0, u_0, y_0)$ , if the starting point  $w_1$  is chosen sufficiently close to  $w_0$  (see [2], Theorem 5.1.). Let  $\nu$  be defined by Theorem 6.1. Denote by  $B_r(w_0)$  the open ball around  $w_0$  with radius  $r$  in the sense of  $C \times L_p \times L_p$ .

**Theorem 6.2.** Suppose that Assumptions (1.4) and (SSC) are satisfied. Choose  $p > n + 1$  and let  $\gamma > 0$  be such that  $\delta := \nu\gamma < 1$ , and  $B_{\gamma\delta}(w_0) \subset N_\gamma(w_0)$ . Then for any starting point  $w_1 \in B_{\gamma\delta}(w_0)$  the SQP method computes a unique sequence  $w_k$  with

$$\|w_k - w_0\|_{\infty, p} \leq \gamma \delta^{2^k - 1},$$

and  $v_k \in B_{\gamma\delta}(u_0)$  for  $k \geq 2$ .

The proof is identical to that given in [2].

Thus we have local quadratic convergence of the SQP method in  $(x, u, y)$ . More precisely, Theorem 6.2 expresses r-quadratic convergence, while Theorem 6.1 shows q-quadratic convergence of the method.

## References

- Alt, W.: The Lagrange-Newton method for infinite dimensional optimization problems. Numer. Funct. Anal. and Optimization 11 (1990), 201-224.
- Alt, W.: Sequential quadratic programming in Banach spaces. In W. Gotth and D. Pallaschke, editors, Advances in Optimization, number 382 in Lecture Notes in Economics and Mathematical Systems, 281-301, Springer-Verlag, 1992.
- Alt, W. and K. Malanowski: The Lagrange-Newton method for nonlinear optimal control problems. Computational Optimization and Applications, 2 (1993), 77-100.
- Alt, W., Sontag, R., and F. Tröltzsch: An SQP method for optimal control of weakly singular Hammerstein integral equations. DFG-Schwerpunktprogramm "Anwendungsbezogene Optimierung und Steuerung", Report No. 423 (1993), to appear in Appl. Math. Optimization.
- Amann, H.: Parabolic evolution equations with nonlinear boundary conditions. In F.E. Browder, editor, Nonlinear Functional Analysis Proc. Sympos. Pure Math., Vol. 35, Part 1, 17-27, 1985.
- Goldberg, H. and F. Tröltzsch: Second order optimality conditions for nonlinear parabolic boundary control problems. SIAM J. Contr. Opt. 31 (1993), 1007-1025.
- Kelley, C.T. and S.J. Wright: Sequential quadratic programming for certain parameter identification problems. Math. Programming 51 (1991), 281-305.
- Kupfer, S.-F. and Sachs, E.W.: Numerical solution of a nonlinear parabolic control problem by a reduced sqp method. Computational Optimization and Applications 1 (1992), 113-135.
- Krasnosel'kii, M.A. et al.: Linear operators in spaces of summable functions (in Russian). Nauka, Moscow, 1966.
- Levitin, E.S. and B.T. Polyak: Constrained minimization methods. USSR J. Comput. Math. and Math. Phys. 6 (1966), 1-50.

11. Tröltzsch, F.: On the semigroup approach for the optimal control of semilinear parabolic equations including distributed an boundary control. *Z. für Analysis und Anwendungen* 8 (1989), 431-443.
12. Tröltzsch, F.: An SQP Method for Optimal Control of a Nonlinear Heat Equation (1993), to appear in *Control Cybernetics*.

Fredi Tröltzsch  
Department of Mathematics  
Technical University of Chemnitz-Zwickau  
PSF 964  
09009 Chemnitz, Germany



# CONDITIONAL STABILITY IN DETERMINATION OF DENSITIES OF HEAT SOURCES IN A BOUNDED DOMAIN

MASAHIRO YAMAMOTO

Department of Mathematical Sciences  
University of Tokyo

ABSTRACT. We consider the heat equation in a bounded domain  $\Omega \subset \mathbb{R}^r$ :

$$\begin{aligned} \frac{\partial u}{\partial t}(x, t) &= \Delta u(x, t) + \sigma(t)f(x) \quad (x \in \Omega, 0 < t < T), \\ u(x, 0) &= 0 \quad (x \in \Omega), \quad \frac{\partial u}{\partial n}(x, t) = 0 \quad (x \in \partial\Omega, 0 < t < T). \end{aligned}$$

Assuming that  $\sigma$  is a known function with  $\sigma(0) \neq 0$ , we prove: (1)  $f(x)$  ( $x \in \Omega$ ) can be uniquely determined from the boundary data  $u(x, t)$  ( $x \in \partial\Omega, 0 < t < T$ ). (2) If  $f$  is restricted to a compact set in a Sobolev space, then we get an estimate:

$$\|f\|_{L^2(\Omega)} = O\left(\left(\log \frac{1}{\eta}\right)^{-\beta}\right) \quad \text{as } \eta \equiv \|u(\cdot, \cdot)\|_{H^1(0, T; L^2(\partial\Omega))} \downarrow 0.$$

Here the exponent  $\beta$  is given by the order of the Sobolev space which is assumed to contain the set of  $f$ 's.

19 *Mathematics Subject Classification.* 35K05, 35R25, 35R30, 93B30

*Key words and phrases.* Conditional stability, a priori bound, biorthogonal function, boundary observation, density of heat source.

## 1. Introduction

We consider an initial/boundary value problem for the heat flow equation:

$$(1.1) \quad \left\{ \begin{array}{l} \frac{\partial u}{\partial t}(x, t) = \Delta u(x, t) + \sigma(t)f(x) \quad (x \in \Omega, 0 < t < T) \\ u(x, 0) = 0 \quad (x \in \Omega) \\ \frac{\partial u}{\partial n}(x, t) = 0 \quad (x \in \partial\Omega, 0 < t < T). \end{array} \right.$$

Here  $\Omega \subset \mathbb{R}^r$  is a bounded domain with smooth boundary  $\partial\Omega$ ,  $\Delta$  is the Laplacian and  $\frac{\partial}{\partial n}$  is a trace operator (e.g. Adams [1]). In particular, if  $u$  is sufficiently smooth on  $\bar{\Omega}$ , then  $\frac{\partial u}{\partial n}(x) = \sum_{i=1}^r \nu_i(x) \frac{\partial u}{\partial x_i}(x)$  ( $x \in \partial\Omega$ ), where  $\nu(x) = (\nu_1(x), \dots, \nu_r(x))$

is the outward unit normal to  $\partial\Omega$  at  $x$ . Throughout this paper,  $T > 0$  is arbitrarily fixed.

The term  $\sigma(t)f(x)$  is considered a heat source. We assume that  $\sigma$  is a known non-zero  $C^1$ -function and is independent of the space variable  $x$ , and  $f \in L^2(\Omega)$  is unknown to be determined from boundary data:

$$(1.2) \quad u(x, t) \equiv h(x, t) \quad (x \in \partial\Omega, 0 < t < T).$$

In (1.2), we regard  $u(x, t)$  ( $x \in \partial\Omega, 0 < t < T$ ) as a trace of  $u(\cdot, t)$  to the boundary. In fact, since  $f \in L^2(\Omega)$  and  $\sigma \in C^1[0, T]$ , there exists a unique *strong solution*  $u \in C^0([0, T]; L^2(\Omega)) \cap C^1((0, T]; L^2(\Omega))$  such that  $\Delta u \in C^0((0, T]; L^2(\Omega))$  and  $u(\cdot, t) \in X_0$  ( $0 < t \leq T$ ) (e.g. Pazy [13]). Here we set  $X_0 = \{u \in H^2(\Omega); \frac{\partial u}{\partial n} = 0\}$  where  $H^2(\Omega) = W^{2,2}(\Omega)$  is a Sobolev space (e.g. [1]). Therefore the left hand side of (1.2) can be considered as the trace.

In this paper, we consider

**Inverse Problem.** (I) (Uniqueness) Do the boundary data  $h(x, t) = u(x, t)$  ( $x \in \partial\Omega, 0 < t < T$ ) determine  $f$  uniquely? (II) (Stability) Establish an actual stability estimate for the correspondence  $h \longrightarrow f$ .

This inverse problem is related to determining the density of radioactive heat sources by the thermal radiation on the surface, for example, if  $\sigma(t) = \exp(-\lambda_0 t)$  (e.g. Lavrent'ev, Romanov and Vasil'ev [11, pp.49-50], Romanov [14, pp.208-210]).

To the uniqueness, we can get the positive answer (Theorem 1 in §2).

To the stability property, it seems that we cannot generally expect any stability in usual norms such as  $\|f\|_{L^2(\Omega)}$  and  $\|h(\cdot, \cdot)\|_{H^1(0, T; L^2(\partial\Omega))}$ . This can be conjectured by the regularity property of the parabolic equation (e.g. [13]). Here and henceforth we set

$$\|h(\cdot, \cdot)\|_{H^1(0, T; L^2(\partial\Omega))} = \left( \int_0^T \int_{\partial\Omega} \left( |h(x, t)|^2 + \left| \frac{\partial h}{\partial t}(x, t) \right|^2 \right) dS_x dt \right)^{\frac{1}{2}}.$$

However, since the uniqueness result can be established and the map  $f \longrightarrow h$  is proved to be continuous from  $L^2(\Omega)$  to  $H^1(0, T; L^2(\partial\Omega))$ , by Tikhonov's theorem (e.g. Lavrent'ev, Romanov and Shishat'skii [10, p.28] or [14, §1.2]), we can see the continuity of the map  $h \longrightarrow f$  provided that we restrict an admissible set of  $f$  to any compact set  $\mathcal{U}$  in  $L^2(\Omega)$ . That is, for any  $\epsilon > 0$ , there exists some  $\delta(\epsilon) > 0$  such that

$$\|u(f) - u(g)\|_{H^1(0, T; L^2(\partial\Omega))} < \delta(\epsilon) \quad \text{and} \quad g, h \in \mathcal{U}$$

imply  $\|g - h\|_{L^2(\Omega)} \leq \epsilon$ . This is what is called a conditional stability. Here  $\delta(\epsilon)$  depends also on the choice of a compact set  $\mathcal{U}$  and, in general, we cannot specify the order of  $\delta(\epsilon)$ .

The purpose of this paper is to show the uniqueness in our inverse problem and to give the modulus  $\delta(\epsilon)$  of continuity if we take a bounded set in some Sobolev space as an admissible set of unknown densities  $f$ 's.

For determination of spatially non-homogeneous terms in parabolic equations, we can refer, for example, to Cannon [3], Cannon and Esteve [4], [11], [14] and Yamamoto [18].

This paper is composed of four sections and an appendix. In §2 we will state Theorem 1 (uniqueness) and Theorem 2 (conditional stability) as our main results. In §3 and §4, we will prove them. The appendix is devoted to the proof of another uniqueness result in the case where the observation time length  $T$  is infinite.

## 2. Main Results

Henceforth we assume

$$(2.1) \quad \sigma(0) \neq 0, \quad \sigma \in C^1[0, T].$$

Let  $L^2(\Omega)$  be the space of real-valued square integrable functions with the norm  $\|\cdot\|$  and the inner product  $(\cdot, \cdot)$ .

For stating our results, we introduce some notations and definitions. Let  $A$  be the realization of  $-\Delta$  in  $L^2(\Omega)$  with Neumann boundary condition:  $Au(x) = -\Delta u(x)$  and  $\mathcal{D}(A) = \{u \in H^2(\Omega); \frac{\partial u}{\partial n}|_{\partial\Omega} = 0\}$ . Let us number the eigenvalues of  $A$  repeatedly according to their multiplicities:

$$0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots,$$

That is, if the multiplicity of  $\lambda_i$  is  $m_i$ , then  $\lambda_i$  appears in the above sequence  $m_i$  times. Let  $\phi_k$  be an eigenfunction for the eigenvalue  $\lambda_k$  of  $A$  ( $k \geq 1$ ). We can choose  $\{\phi_k\}_{k \geq 1}$  such that

$$(2.2) \quad (\phi_k, \phi_l) = \delta_{kl}.$$

Here we set  $\delta_{kl} = 1$  ( $k = l$ ),  $= 0$  (otherwise).

Let us fix  $\alpha > 0$  such that

$$(2.3) \quad \alpha > \frac{r-1}{2}.$$

For any  $M > 0$ , let us define an admissible set of unknown terms  $f$ 's by

$$(2.4) \quad \mathcal{U}_{M,\alpha} = \{f \in L^2(\Omega); \sum_{k=1}^{\infty} |(f, \phi_k)|^2 k^{\frac{2\alpha}{r}} \leq M^2\}.$$

For  $\frac{\alpha}{2} \in \mathbb{N}$ , by using the asymptotic behaviour of  $\lambda_n$  (see (3.1) below), we can replace the definition of  $\mathcal{U}_{M,\alpha}$  by

$$(2.5) \quad \mathcal{U}_{M,\alpha} = \{f \in H^\alpha(\Omega); \frac{\partial}{\partial n}(A^j f) = 0, 0 \leq j \leq \frac{\alpha}{2} - 1, \|f\|_{H^\alpha(\Omega)} \leq M\}.$$

**Remark 1.** We can define the fractional power  $(A + \gamma)^\beta$  for  $\gamma > 0$  and  $\beta \in \mathbb{R}$  (e.g. [13]). By  $(A + \gamma)^{\frac{\alpha}{2}}$ , we can rewrite the definition of  $\mathcal{U}_{M,\alpha}$ :

$$\mathcal{U}_{M,\alpha} = \{ f \in L^2(\Omega); f \in \mathcal{D}\left((A + \gamma)^{\frac{\alpha}{2}}\right), \| (A + \gamma)^{\frac{\alpha}{2}} f \| \leq M \}.$$

Now we are ready to state

**Theorem 1 (Uniqueness):** *Let  $f$  satisfy*

$$\sum_{k=1}^{\infty} |(f, \phi_k)|^2 k^{\frac{2\alpha}{r}} < \infty.$$

*Let  $u(f)(x, t)$  be the strong solution to*

$$(2.6) \quad \left\{ \begin{array}{l} \frac{\partial u}{\partial t}(x, t) = \Delta u(x, t) + \sigma(t)f(x) \quad (x \in \Omega, 0 < t < T) \\ u(x, 0) = 0 \quad (x \in \Omega) \\ \frac{\partial u}{\partial n}(x, t) = 0 \quad (x \in \partial\Omega, 0 < t < T). \end{array} \right.$$

*If*

$$(2.7) \quad u(f)(x, t) = 0 \quad (x \in \partial\Omega, 0 < t < T),$$

*then  $f(x) = 0$  for almost all  $x \in \Omega$ .*

**Theorem 2 (Conditional Stability).** *For any fixed  $\alpha > \frac{r-1}{2}$  and any  $M > 0$ , we a priori assume that  $f \in \mathcal{U}_{M,\alpha}$ . If (2.6) holds, then there exists a constant  $C > 0$  such that*

$$(2.8) \quad \|f\| \leq \frac{M}{N^{\frac{\alpha}{r}}} + C \exp(CN^{\frac{1}{r}}) \|u(f)\|_{H^1(0,T;L^2(\partial\Omega))}$$

*for any  $N \in \mathbb{N}$ . Moreover, as  $\eta \equiv \|u(f)\|_{H^1(0,T;L^2(\partial\Omega))}$  tends to zero, we have the estimate:*

$$(2.9) \quad \|f\| = O\left(\left(\frac{1}{\log \frac{1}{\eta}}\right)^\beta\right) \quad \text{for any } \beta < \alpha.$$

**Remark 2.** By (4.3) in the proof of Theorem 2 in §4, we can see: As the norm for densities  $f$ 's, if we adopt  $\|f\| = \sup_{k \geq 1} |(f, \phi_k)| \exp(-C_1 k^{\frac{1}{r}})$  for a sufficiently large  $C_1 > 0$  and we define an admissible set of  $f$ 's by  $\tilde{X} = \{f \in L^2(\Omega); \|f\| < \infty\}$ , then the map  $u(x, t) \ (x \in \partial\Omega, 0 < t < T) \rightarrow f(x) \ (x \in \Omega)$  is Lipschitz continuous from  $H^1(0, T; L^2(\partial\Omega))$  to  $\tilde{X}$ . However, the topology in  $\tilde{X}$  is too weak in the sense that we can not generally get the continuity of the map  $f \rightarrow u(f)$  from  $\tilde{X}$  to  $H^1(0, T; L^2(\partial\Omega))$ .

In [3], a similar problem is considered in the case of  $\sigma(i) \equiv -1$  and the maximum principle for the analytic function is a key. On the other hand, in this paper we reduce our inverse problem to some "moment problem" by which we can estimate Fourier coefficients of  $f$ . Our technique is control-theoretical, and is derived from Russell [15], where a boundary control problem is considered for heat equation.

If  $\Omega$  is an  $r$ -dimensional parallelepipedon  $\{(x_1, \dots, x_r); 0 < x_i < L_i \ (1 \leq i \leq r)\}$ , ( $L_i > 0 : 1 \leq i \leq r$ ), then we can reach similar results for uniqueness and conditional stability even though we restrict observations on the whole boundary,  $u(x, t) \ (x \in \partial\Omega, 0 < t < T)$  to a part of  $\partial\Omega$  ([18]). The method in [18] is based on Vattorini [6]. However for a general bounded domain, it is not certain whether or not boundary data restricted to a part of  $\partial\Omega$  can guarantee conditional stability such as Theorem 2. As is mentioned in §1, in [11] and [14], a similar problem is considered in the case of  $\Omega = \{(x_1, \dots, x_{r-1}, x_r); -\infty < x_1, \dots, x_{r-1} < \infty, x_r > 0\}$  and  $T = \infty$ , where the key technique is the Laplace transform with respect to time  $t$  and so it seems essential that the observation is over any  $t > 0$ , not over a finite time interval  $0 < t < T (< \infty)$ .

**Remark 3.** For discussing parabolic inverse problems, another method by Reznitskaja is applicable (e.g. Lavrent'ev, Reznitskaja and Yakhno [9, pp.6-8], and Romanov [14, pp.213-218]). That is, if we can observe boundary values over  $[0, \infty)$  of the  $t$ -axis, then we can replace our inverse problem for a parabolic equation (1.1) by a determination problem of  $f$  in a hyperbolic equation  $\frac{\partial^2 u}{\partial t^2} = \Delta u + \sigma(t)f(x) \ (x \in \Omega, t > 0)$ . Then we can apply results in hyperbolic inverse problems (e.g. [10, Chapter VII] and [14, Chapter 4]). In our paper, however, by constructing biorthogonal functions, we directly discuss a parabolic inverse problem.

If we assume that  $T = \infty$ , then we can weaken the condition (2.7) for the uniqueness. That is,

**Proposition (Uniqueness).** *Let us assume that  $f$  and  $\sigma$  are continuous and bounded respectively in  $\Omega$  and  $[0, \infty)$  and that  $\sigma$  does not vanish identically. If  $u(f)(x, t)$  satisfies*

$$(2.10) \quad \left\{ \begin{array}{l} \frac{\partial u}{\partial t}(x, t) = \Delta u(x, t) + \sigma(t)f(x) \quad (x \in \Omega, t > 0) \\ u(x, 0) = 0 \quad (x \in \Omega) \\ \frac{\partial u}{\partial n}(x, t) = 0 \quad (x \in \partial\Omega, t > 0), \end{array} \right.$$

and

$$(2.11) \quad u(f)(x, t) = 0 \quad (x \in \text{any open set of } \partial\Omega, t > 0),$$

then  $f(x) = 0 \ (x \in \Omega)$ .

We can prove a uniqueness result of this type in the case where  $\Delta$  in (2.10) is replaced by any uniformly elliptic differential operator of the second order, but we will omit it. For convenience, the proof of this proposition is given in Appendix.

### 3. Proof of Theorem 1

We will divide the proof into the following three steps.

1) In this step, we will prove that the trace of the series at the boundary

$$\sum_{k=1}^{\infty} \left( \int_0^t \exp(-\lambda_k(t-s)) \sigma(s) ds \right) (f, \phi_k) \phi_k(x)$$

is convergent in  $H^1(0, T; L^2(\partial\Omega))$ . First, by a theorem of asymptotic distribution of eigenvalues (e.g. Agmon [2], Courant and Hilbert [5]), we have

$$(3.1) \quad \lambda_k = c_2 k^{\frac{2}{d}} + o(k^{\frac{2}{d}})$$

as  $k \rightarrow \infty$ , where  $c_2$  is a positive constant.

For simplicity, we set

$$(3.2) \quad \psi_k(t) = \int_0^t \exp(-\lambda_k(t-s)) \sigma(s) ds \quad (k \geq 1, 0 \leq t \leq T).$$

By  $\sigma \in C^1[0, T]$ , we have

$$|\psi_k(t)| = \left| \int_0^t \exp(-\lambda_k s) \sigma(t-s) ds \right| \leq C_3' \int_0^t \exp(-\lambda_k s) ds$$

and

$$\left| \frac{d\psi_k}{dt}(t) \right| \leq C_3' \left( \exp(-\lambda_k t) + \int_0^t \exp(-\lambda_k s) ds \right) \quad (k \geq 1).$$

Therefore we get

$$(3.3) \quad \|\psi_k\|_{H^1(0, T)} \leq \frac{C_3'}{\sqrt{\lambda_k + 1}} \quad (k \geq 1).$$

On the other hand, we have  $\mathcal{D}((A+1)^{\frac{1}{4}}) = H^{\frac{1}{2}}(\Omega)$  and  $\|u\|_{H^{1/2}(\Omega)} \leq C_3'' \|(A+1)^{\frac{1}{4}} u\|$  (e.g. Fujiwara [7]) and  $\|u\|_{L^2(\partial\Omega)} \leq C_3'' \|u\|_{H^{1/2}(\Omega)}$  (e.g. Adams [1]). Consequently by (2.2) we see

$$\|\phi_k\|_{L^2(\partial\Omega)} \leq C_3'^2 \|(A+1)^{\frac{1}{4}} \phi_k\| = C_3'^2 \|(\lambda_k + 1)^{\frac{1}{4}} \phi_k\| = C_3'^2 (\lambda_k + 1)^{\frac{1}{4}}.$$

so that we have

$$\begin{aligned}
 & \sum_{k=1}^{\infty} \left\| \left( \int_0^t \exp(-\lambda_k(t-s)) \sigma(s) ds \right) (f, \phi_k) \phi_k \right\|_{H^1(0,T;L^2(\partial\Omega))} \\
 &= \sum_{k=1}^{\infty} \|(f, \phi_k) \psi_k \phi_k\|_{H^1(0,T;L^2(\partial\Omega))} \\
 &\leq \sum_{k=1}^{\infty} \frac{C_3}{(\lambda_k + 1)^{1/2}} \times |(f, \phi_k)| \times C_3^2 (\lambda_k + 1)^{\frac{1}{4}} \quad (\text{by (3.3)}) \\
 &= C_3 \sum_{k=1}^{\infty} |(f, \phi_k)| (\lambda_k + 1)^{-\frac{1}{4}} \\
 &= C_3 \sum_{k=1}^{\infty} |(f, \phi_k)| (\lambda_k + 1)^{\frac{\alpha}{2}} (\lambda_k + 1)^{-\frac{\alpha}{2} - \frac{1}{4}} \\
 &\leq C_3 \left( \sum_{k=1}^{\infty} |(f, \phi_k)|^2 (\lambda_k + 1)^{\alpha} \right)^{\frac{1}{2}} \left( \sum_{k=1}^{\infty} (\lambda_k + 1)^{-\alpha - \frac{1}{2}} \right)^{\frac{1}{2}} \\
 &\quad (\text{by Schwarz's inequality}) \\
 &\leq C_3'' \left( \sum_{k=1}^{\infty} |(f, \phi_k)|^2 k^{\frac{2\alpha}{r}} \right)^{\frac{1}{2}} \left( \sum_{k=1}^{\infty} k^{-\frac{2\alpha+1}{r}} \right)^{\frac{1}{2}} \quad (\text{by (3.1)}) \\
 &< \infty \quad (\text{by } f \in \mathcal{U}_{M,\alpha} \text{ and } \alpha > \frac{r-1}{2}).
 \end{aligned}$$

Thus, by Itô [8] for almost all  $x \in \partial\Omega$  and almost all  $t \in [0, T]$ , we can obtain

$$(3.4) \quad u(f)(x, t) = \sum_{k=1}^{\infty} (f, \phi_k) \psi_k(t) \phi_k(x)$$

and we see that the series is convergent in  $H^1(0, T; L^2(\partial\Omega))$ .

2) In this step, on the basis of Russell [15], in  $H^1(0, T; L^2(\partial\Omega))$  we will construct a biorthogonal system  $\{\theta_k(x, t)\}_{k \geq 1}$  to  $\{\psi_k(t) \phi_k(x)\}_{k \geq 1}$ .

By [15], there exists a system  $\{\eta_k(x, t)\}_{k \geq 1}$  in  $L^2(0, T; L^2(\partial\Omega))$  such that

$$(3.5) \quad \langle \exp(-\lambda_k t) \phi_k, \eta_l \rangle_{L^2} = \delta_{kl} \quad (k, l \geq 1).$$

Here and henceforth we denote the inner products in  $L^2(0, T; L^2(\partial\Omega))$  and in  $H^1(0, T; L^2(\partial\Omega))$  respectively by  $\langle \cdot, \cdot \rangle_{L^2}$  and  $((\cdot, \cdot))_{H^1}$ :

$$\langle g, h \rangle_{L^2} = \int_0^T \int_{\partial\Omega} g(x, t) h(x, t) dS_x dt$$

and

$$((g, h))_{H^1} = \int_0^T \int_{\partial\Omega} \left( g(x, t)h(x, t) + \frac{\partial g}{\partial t}(x, t) \frac{\partial h}{\partial t}(x, t) \right) dS_x dt.$$

Let us define an operator  $K : L^2(0, T; L^2(\partial\Omega)) \longrightarrow H^1(0, T; L^2(\partial\Omega))$  by

$$(Kg)(x, t) = \int_0^t \sigma(t-s)g(x, s)ds \quad (0 < t < T, x \in \partial\Omega).$$

We can easily see that  $K$  is bounded from  $L^2(0, T; L^2(\partial\Omega))$  to  $H^1(0, T; L^2(\partial\Omega))$ . By the assumption (2.1) on  $\sigma$ , we have  $\sigma(0) \neq 0$ , so that we can prove that there exists a constant  $C_4 > 0$  such that

$$(3.6) \quad C_4 \|Kg\|_{H^1(0, T; L^2(\partial\Omega))} \geq \|g\|_{L^2(0, T; L^2(\partial\Omega))}$$

for any  $g \in L^2(0, T; L^2(\partial\Omega))$ . In fact, let

$$h(x, t) = (Kg)(x, t) = \int_0^t \sigma(t-s)g(x, s)ds \quad (0 < t < T, x \in \partial\Omega).$$

Then we have

$$\frac{\partial h}{\partial t}(x, t) = \sigma(0)g(x, t) + \int_0^t \frac{d\sigma}{dt}(t-s)g(x, s)ds \quad (0 < t < T, x \in \partial\Omega),$$

which is a Volterra integral equation in  $t$  of the second kind, so that we can construct the real-valued resolvent kernel  $\Gamma(t, s)$  for  $\frac{1}{\sigma(0)} \frac{d\sigma}{dt}(t-s)$  ( $0 \leq s \leq t \leq T$ ) and we have

$$g(x, t) = \frac{1}{\sigma(0)} \frac{\partial h}{\partial t}(x, t) - \frac{1}{\sigma(0)} \int_0^t \Gamma(t, s) \frac{\partial h}{\partial t}(x, s)ds$$

(e.g. Tricomi [17]). This means (3.6).

Now (3.6) implies that the range of  $K$  is closed in  $H^1(0, T; L^2(\partial\Omega))$ . Therefore we see that  $(K^{-1})^*$  exists and  $(K^{-1})^* = (K^*)^{-1}$ . Here and henceforth we identify the adjoint space  $X^*$  of a Hilbert space with  $X$  and  $\cdot^*$  denotes the adjoint of an operator under consideration.

We define a system  $\{\theta_l\}_{l \geq 1} \subset H^1(0, T; L^2(\partial\Omega))$  by

$$(3.7) \quad \theta_l(x, t) = ((K^{-1})^* \eta_l)(x, t) \quad (l \geq 1, x \in \partial\Omega, 0 < t < T)$$

This system  $\{\theta_l\}_{l \geq 1}$  is the desired one, that is,

$$(3.8) \quad ((\psi_k \phi_k, \theta_l))_{H^1} = \delta_{kl} \quad (k \geq 1, l \geq 1).$$

Here let us recall that  $\psi_k$  is given by (3.2). The relation (3.8) can be proved as follows:

$$\begin{aligned} ((\psi_k \phi_k, \theta_l))_{H^1} &= ((K(\exp(-\lambda_k t)\phi_k), (K^{-1})^* \eta_l))_{H^1} \\ &= \langle (K^{-1})^{**} K(\exp(-\lambda_k t)\phi_k), \eta_l \rangle_{L^2} \\ &= \delta_{kl}, \end{aligned}$$



by (3.5) and (3.7).

3) We will complete the proof of Theorem 1. By the assumption, we have

$$u(f)(x, t) = 0 \quad (x \in \partial\Omega, 0 < t < T).$$

Therefore (3.4) implies

$$\sum_{k=1}^{\infty} (f, \phi_k) \psi_k(t) \phi_k(x) = 0 \quad \text{in } H^1(0, T; L^2(\partial\Omega)).$$

Since, as is proved in the first step 1), this series is convergent in  $H^1(0, T; L^2(\partial\Omega))$ , we can take the  $H^1(0, T; L^2(\partial\Omega))$ -inner product with  $\theta_l$  in termwise, so that

$$(f, \phi_l) = 0 \quad (l \geq 1)$$

by the biorthogonality (3.8). Thus the proof is complete.

#### 4. Proof of Theorem 2

First we will estimate norms of  $\theta_k$  constructed by (3.7) in  $H^1(0, T; L^2(\partial\Omega))$ . By Russell [15], there exists a constant  $C'_5 > 0$  such that

$$\|\eta_l\|_{L^2(0, T; L^2(\partial\Omega))} \leq C'_5 \exp(C'_5 \sqrt{\lambda_l}) \quad (l \geq 1).$$

Consequently by (3.1) we can obtain

$$(4.1) \quad \|\eta_l\|_{L^2(0, T; L^2(\partial\Omega))} \leq C'_5 \exp(C'_5 l^{\frac{1}{2}}) \quad (l \geq 1).$$

By (3.6) and (3.7), we have

$$(4.2) \quad \begin{aligned} \|\theta_l\|_{H^1(0, T; L^2(\partial\Omega))} &\leq \|(K^{-1})^*\| \times \|\eta_l\|_{L^2(0, T; L^2(\partial\Omega))} \\ &\leq C'_5 C'_5 \exp(C'_5 l^{\frac{1}{2}}) \equiv C'_6 \exp(C'_6 l^{\frac{1}{2}}) \quad (l \geq 1). \end{aligned}$$

Here  $\|(K^{-1})^*\|$  is the operator norm of  $(K^{-1})^*$  and  $C'_6 > 0$  is a constant which is independent of  $l \geq 1$ .

Next we will estimate the Fourier coefficients  $(f, \phi_k)$  by using (4.2) and (3.4), and complete the proof of Theorem 2. Since the series in (3.4) converges in  $H^1(0, T; L^2(\partial\Omega))$ , we can take the inner product  $((\cdot, \cdot))_{H^1}$  with  $\theta_l$  in termwise. Consequently by (3.8) we can reach

$$((u(f), \theta_l))_{H^1} = (f, \phi_l) \quad (l \geq 1),$$

and so by (4.2) we obtain

$$(4.3) \quad |(f, \phi_l)| \leq C'_6 \|u(f)\|_{H^1(0, T; L^2(\partial\Omega))} \exp(C'_6 l^{\frac{1}{2}}) \quad (l \geq 1).$$

First let us prove (2.8). We have

$$\begin{aligned}
 \|f\|^2 &= \sum_{l=1}^{\infty} |(f, \phi_l)|^2 = \sum_{l=1}^{N-1} |(f, \phi_l)|^2 + \sum_{l=N}^{\infty} \frac{|(f, \phi_l)|^2 l^{\frac{2\alpha}{r}}}{l^{\frac{2\alpha}{r}}} \\
 &\leq \sum_{l=1}^{N-1} |(f, \phi_l)|^2 + N^{-\frac{2\alpha}{r}} M^2 \quad (\text{by } f \in \mathcal{U}_{M,\alpha}) \\
 &\leq C_6^2 \|a(f)\|_{H^1((0,T);L^2(\partial\Omega))}^2 \sum_{l=1}^{N-1} \exp(2C_6 l^{\frac{1}{r}}) + N^{-\frac{2\alpha}{r}} M^2 \quad (\text{by (1.3)}) \\
 &\leq C_6^2 \|a(f)\|_{H^1((0,T);L^2(\partial\Omega))}^2 (N-1) \exp(2C_6(N-1)^{\frac{1}{r}}) + N^{-\frac{2\alpha}{r}} M^2
 \end{aligned}$$

which is (2.8).

Next let us consider (2.9). For  $\beta < \alpha$ , we fix the natural number  $N$  such that

$$M^{\frac{1}{\alpha}} \left( \log \frac{1}{\eta} \right)^{\frac{1}{\alpha}} \leq N < 1 + M^{\frac{1}{\alpha}} \left( \log \frac{1}{\eta} \right)^{\frac{1}{\alpha}}.$$

Then we can get

$$(4.4) \quad \frac{M}{N^{\frac{\beta}{r}}} \leq \left( \frac{1}{\log \frac{1}{\eta}} \right)^{\beta}.$$

Moreover we have

$$\begin{aligned}
 C\eta \exp(CN^{\frac{1}{r}}) &\leq C\eta \exp \left( C \left( 1 + M^{\frac{1}{\alpha}} \left( \log \frac{1}{\eta} \right)^{\frac{1}{\alpha}} \right)^{\frac{1}{r}} \right) \\
 &\leq C'\eta \exp \left( CM^{\frac{1}{\alpha}} \left( \log \frac{1}{\eta} \right)^{\frac{1}{\alpha}} \right).
 \end{aligned}$$

Therefore, in order to prove (2.9), by (2.8) and (4.4) it is sufficient to verify that

$$\eta \exp \left( CM^{\frac{1}{\alpha}} \left( \log \frac{1}{\eta} \right)^{\frac{1}{\alpha}} \right) = O \left( \left( \frac{1}{\log \frac{1}{\eta}} \right)^{\beta} \right)$$

as  $\eta \downarrow 0$  for  $\beta < \alpha$ , which is seen from

$$\lim_{\eta \downarrow 0} \eta \exp \left( CM^{\frac{1}{\alpha}} \left( \log \frac{1}{\eta} \right)^{\frac{1}{\alpha}} \right) \left( \log \frac{1}{\eta} \right)^{\beta} = 0 \quad (\beta < \alpha).$$

### Appendix. Proof of Proposition

Let  $U(t, x, y)$  be the fundamental solution of  $\frac{\partial}{\partial t} - \Delta$  with the zero Neumann boundary condition. Then under the assumption on regularity of  $\sigma$  and  $f$ , we can express the solution to (2.10) as

$$u(f)(x, t) = \int_0^t \sigma(s) \left( \int_{\Omega} U(t-s, x, y) f(y) dy \right) ds \quad (x \in \Omega, t > 0)$$

(Itô [8]). We set

$$u_x(t) = \int_{\Omega} U(t, x, y) f(y) dy$$

for any fixed  $x \in \bar{\Omega}$ . Let us assume that  $u(f)(x, t) = 0$  for  $t > 0$  and any  $x \in \Gamma$ ; an open set of  $\partial\Omega$ . Therefore we have

$$\int_0^t \sigma(s) u_x(t-s) ds = 0 \quad (t > 0, x \in \Gamma),$$

and since  $\sigma$  does not vanish identically and  $\sigma, u_x(\cdot)$  are continuous, by Titchmarsh's theorem on the convolution (e.g. Titchmarsh [16, p.322]), we can obtain

$$u_x(t) \equiv \int_{\Omega} U(t, x, y) f(y) dy = 0 \quad (t > 0, x \in \Gamma).$$

This implies that for any  $0 < \epsilon < T$ , the classical solution  $u$  to

$$\left\{ \begin{array}{ll} \frac{\partial u}{\partial t}(x, t) = \Delta u(x, t) & (x \in \Omega, \epsilon < t < T) \\ u(x, 0) = f(x) & (x \in \Omega) \\ \frac{\partial u}{\partial n}(x, t) = 0 & (x \in \partial\Omega, \epsilon < t < T) \end{array} \right\}$$

vanishes in an open set of  $\partial\Omega$ . Therefore it follows from the unique continuation theorem (e.g. Mizohata [12]) that  $u(x, t) = 0$  ( $x \in \Omega, t > \epsilon$ ). Consequently by analyticity of  $u$  in  $t$  (e.g. [13]), we see  $u(x, t) = 0$  ( $x \in \Omega, t \geq 0$ ). Thus we can see  $f(x) = 0$  ( $x \in \Omega$ ).

**Acknowledgements.** The author would like to express his hearty thanks to referees for valuable comments, and sincerely thanks the Alexander von Humboldt Foundation for the financial support. Moreover he expresses his hearty gratitude to Professor Dr. Karl-Heinz Hoffmann of Technische Universität München.

## References

1. R. A. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.
2. S. Agmon, *Lectures on Elliptic Boundary Value Problems*, Van Nostrand, Princeton, N.J., 1965.
3. J. R. Cannon, *Determination of an unknown heat source from overspecified boundary data*, SIAM J. Numer. Anal. **5** (1968), 275–286.
4. J. R. Cannon and S. P. Esteve, *An inverse problem for the heat equation*, Inverse Problems **2** (1986) 395–403.
5. R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Volume I, Interscience, New York, 1953.
6. H. O. Fattorini, *Boundary control of temperature distributions in a parallelepipedon*, SIAM J. Control **13** (1975) 1–13.
7. D. Fujiwara, *Concrete characterization of the domains of fractional powers of some elliptic differential operators of the second order*, Proc. Japan Acad. **43** (1967), 82–86.
8. S. Itô, *Diffusion Equations*, (English translation), American Mathematical Society, Providence, Rhode Island, 1992.
9. M. M. Lavrent'ev, K. G. Reznitskaja and V. G. Yakhno, *One-dimensional Inverse Problems of Mathematical Physics*, (English translation), American Mathematical Society, Providence, Rhode Island, 1986.
10. M. M. Lavrent'ev, V. G. Romanov and S. P. Shishat'skii, *Ill-posed Problems of Mathematical Physics and Analysis*, (English translation), American Mathematical Society, Providence, Rhode Island, 1986.
11. M. M. Lavrent'ev, V. G. Romanov and V. G. Vasil'ev, *Multidimensional Inverse Problems for Differential Equations*, (Lecture Notes in Mathematics, No.167), Springer-Verlag, Berlin, 1970.
12. S. Mizohata, *The Theory of Partial Differential Equations*, Cambridge University Press, London, 1973.
13. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer-Verlag, Berlin, 1983.
14. V. G. Romanov, *Inverse Problems of Mathematical Physics*, (English translation), VNU Science Press, Utrecht, 1987.
15. D. L. Russell, *A unified boundary controllability theory for hyperbolic and parabolic partial differential equations*, Studies in Applied Mathematics **52** (1973), 189–211.
16. E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford University Press, London, 1937.
17. F. G. Tricomi, *Integral Equations*, Dover, New York, 1985.
18. M. Yamamoto, *Conditional stability in determination of force terms of heat equations in a rectangle*, Mathematical and Computer Modelling **18** (1993), 79–88.

Masahiro Yamamoto  
 Department of Mathematical Sciences  
 University of Tokyo  
 3-8-1 Komaba, Meguro, Tokyo 153  
 Japan

E-mail: myama@tansei.cc.u-tokyo.ac.jp

# BOUNDARY STABILIZATION OF THE KORTEWEG-DE VRIES EQUATION

BING-YU ZHANG

Department of Mathematical Sciences  
University of Cincinnati

**ABSTRACT.** Consider herein is the initial-boundary value problem of the KdV equation posed on the bounded interval  $(0, 1)$ :

$$(*) \quad \begin{cases} u_t + uu_x + u_{xxx} = 0, & u(x, 0) = \phi(x) \\ u(0, t) = 0, & u(1, t) = 0, & u_x(1, t) = \alpha u_x(0, t) \end{cases}$$

where  $|\alpha| < 1$ .

It is shown that (i) the system  $(*)$  is globally well-posed in the space

$$H_{\alpha}^3 := \left\{ \phi \in H^3(0, 1); \quad \phi(0, t) = 0, \quad \phi(1, t) = 0, \quad \phi_x(1, t) = \alpha \phi_x(0, t) \right\}$$

and (ii) if  $\alpha \neq 0$ , then the system  $(*)$  is locally well-posed in the space  $H_0^1(0, 1)$ , but its small amplitude solutions exist globally and decay exponentially to zero as  $t \rightarrow \infty$ .

1991 *Mathematics Subject Classification.* 35Q20, 93D15, 93C20, 93B05

*Key words and phrases.* Korteweg-de Vries equation, stabilization, boundary control.

## 1. Introduction

In this paper we consider a boundary control system described by the Korteweg-de Vries (KdV) equation posed on a bounded interval  $(0, 1)$ :

$$(1.1) \quad \begin{cases} u_t + uu_x + u_{xxx} = 0, & u(x, 0) = \phi(x) \\ u(0, t) = 0, & u(1, t) = 0, & u_x(1, t) = h(t) \end{cases}$$

for  $x \in (0, 1)$ ,  $t \geq 0$ , where the subscripts denote partial derivatives and the boundary value function  $h(t)$  is considered as a control input. The goal is to find an appropriate boundary feedback control law so that the resulting closed loop system is stabilized.

---

The paper was completed while the author held a postdoctoral position in industrial mathematics at the Institute for Mathematics and its Applications at the University of Minnesota.

It is readily verified that, for any smooth solution of (1.1), one has

$$\frac{d}{dt} \int_0^1 u^2(x, t) = u_x^2(1, t) - u_x^2(0, t) = h^2(t) - u_x^2(0, t).$$

This suggests to choose a feedback control law

$$(1.2) \quad h(t) = \alpha u_x(0, t), \quad \text{with } |\alpha| < 1,$$

in order to stabilize the system (1.1). Indeed, by choosing  $h$  as in (1.2), one would see that the identity

$$(1.3) \quad \frac{d}{dt} \int_0^1 u^2(x, t) = -(1 - \alpha^2) u_x^2(0, t)$$

holds for smooth solutions of the system (1.1). Hence the "energy"  $\int_0^1 u^2(x, t) dx$  is decreasing as long as  $u_x(0, t)$  does not vanish. Plugging (1.2) into (1.1) one obtains a closed-loop system

$$(1.4) \quad \begin{cases} u_t + uu_x + u_{xxx} = 0, & u(x, 0) = \phi(x), \\ u(0, t) = 0, & u(1, t) = 0, \quad u_x(1, t) = \alpha u_x(0, t). \end{cases}$$

According to (1.3), it possesses a dissipative mechanism introduced through the boundary. Our main concern is whether the introduced dissipative mechanism is strong enough to stabilize the system.

A similar problem has been considered in [16] for the KdV equation posed on the periodic domain:

$$(1.5) \quad \begin{cases} v_t + vv_x + v_{xxx} = 0, & v(x, 0) = \phi(x), \\ v(1, t) = v(0, t), & v_{xx}(1, t) = v_{xx}(0, t), \quad v_x(1, t) - v_x(0, t) = h(t) \end{cases}$$

where a dissipative mechanism is introduced to (1.5) by choosing

$$h(t) = -k(v_x(1, t) + v_x(0, t)), \quad k > 0.$$

The resulting closed loop system is of the form

$$(1.6) \quad \begin{cases} v_t + vv_x + v_{xxx} = 0, & v(x, 0) = \phi(x), \\ v(0, t) = v(1, t), & v_{xx}(1, t) = v_{xx}(0, t), \quad v_x(1, t) = \alpha v_x(0, t) \end{cases}$$

where  $\alpha = (1 - k)/(1 + k)$ . It was shown in [16] that there exists a  $\delta > 0$  and a  $\beta > 0$  such that for any  $\phi \in H^1(0, 1)$  with  $\|\phi\|_{H^1(0, 1)} \leq \delta$ , the system (1.6) has a unique solution  $v \in L^\infty(0, \infty; H^1(0, 1))$  and

$$\|v(\cdot, t) - \phi\|_{L^2(0, 1)} \leq e^{-\beta t} \|\phi - [\phi]\|_{L^2(0, 1)}$$

for any  $t \geq 0$  where  $[\phi]$  is the mean value of  $\phi(x)$  over the interval  $(0, 1)$ , i.e.,

$$[\phi] = \int_0^1 \phi(x) dx.$$

In this paper we are going to show that a similar result is also true for the system (1.4).

We will first consider the well-posedness problem of the system (1.4), which is an interesting initial-boundary value problem for the KdV equation in its own right. In the past several decades there have been many literatures discussing the initial-boundary value problem for the KdV equation. For a beginning of collection of references, we refer to [2], [8]–[10] and [18] for the equation posed on the whole real line  $R$ , or on a periodic domain and refer to [3] for the equation posed on a half real line  $R^+$ . As for the equation posed on a finite interval we refer to [4]. In particular, the following initial boundary value problem has been considered in [4]

$$(1.7) \quad \begin{cases} u_t + uu_x + u_{xxx} = f, & u(x, 0) = 0 \\ \alpha_1 u_{xx}(0, t) + \alpha_2 u_x(0, t) + \alpha_3 u(0, t) = 0 \\ \beta_1 u_{xx}(1, t) + \beta_2 u_x(1, t) + \beta_3 u(1, t) = 0 \\ \xi_1 u_x(1, t) + \xi_2 u(1, t) = 0 \end{cases}$$

where  $\alpha_i, \beta_j, \xi_j \in R, i = 1, 2, 3; j = 1, 2$  and some assumptions are imposed so that the 'energy' of the solutions of (1.7) is decreasing. It was shown in [4] that (1.7) has a unique solution

$$u \in L^2(0, T; H^3(0, 1)), \quad u_t \in L^\infty(0, T; L^2(0, 1)) \cap L^2(0, T; H^1(0, 1))$$

for  $f, f_t \in L^2(0, T; L^2(0, 1))$ . The result is local in the sense that  $T$  depends on  $f$ . A counter example is given in [4] to show that one cannot expect to have global result for (1.7) in general. We point out here that the boundary conditions of the system (1.4) is not covered by (1.7) except the case that  $\alpha = 0$ . In this paper we will show that the system (1.4) including the case  $\alpha = 0$  is globally well-posed in the space  $H_\alpha^3(0, 1)$  where

$$H_\alpha^3(0, 1) := \{g \in H^3(0, 1) : g(0, t) = 0, \quad g(1, t) = 0, \quad g_x(1, t) = \alpha g_x(0, t)\}.$$

The notion of well-posedness includes existence, uniqueness, persistence property (i.e., the solution describes a continuous curve in the space  $H_\alpha^3$ ) and continuous dependence of the solution upon the initial data. We first establish the local well-posedness result using contraction principle and the smoothing properties of the associated linear system. The global result is then obtained by finding global *a priori* estimates for smooth solutions of (1.4). It is interesting to note that while the  $L^2$  estimate of solutions is easy to establish, the global  $H^1$  and  $H^2$  estimates

of solutions are very difficult to obtain. What we do is to find an  $L^2$  estimate of the time derivative of the solution,  $u_t$  which, in turn, would provide us the needed global  $H^3$  estimate. Our argument relies heavily on the boundary conditions  $u(0, t) = u(1, t) = 0$ . The approach used here does not apply to the system (1.6) considered in [16]. In fact, only local well-posedness result was obtained in [16] for (1.6). It remains open whether the solutions of the system (1.6) exist globally in time or blows up in finite time (see [17]).

After showing the closed loop system (1.4) is well-posed, we show that the dissipative mechanism introduced through the boundary is strong enough to enable the small amplitude solutions of the system (1.4) ( $\alpha \neq 0$ ) decay exponentially. The approach we will use is similar to that used in [16]. Following the lines outlined there, we will begin with studying spectral properties of the operator  $A_\alpha$  defined by

$$(1.8) \quad (A_\alpha g)(x) = -g'''(x)$$

with domain  $\mathcal{D}(A_\alpha) = H_\alpha^3$ . In particular, we show that the operator  $A_\alpha$  is a discrete spectral operator whose eigenfunctions form a Riesz basis for the space  $L^2(0, 1)$  which would provide us a working ground to apply nonharmonic analysis techniques (cf. [14]). Then some smoothing properties of the associated linear system will be presented and will be used to show that the system (1.4) is locally well-posed in the space  $H_0^1(0, 1)$ , but small amplitude solutions exist globally. As in [16] the decay property of small amplitude solutions is obtained by using an infinite dimension version of the second method of Lyapounov. It is not clear how large amplitude solutions of the system (1.4) behave as  $t \rightarrow \infty$ .

The paper is organized as follows. In section 2, we present the global well-posedness results for the system (1.4). In section 3, we study spectral properties of the operator  $A_\alpha$ . The decay property of small amplitude solutions of (1.4) is established in section 4.

## 2. Global Well-posedness

In this section we show that the closed loop system (1.4) is globally well-posed in the space  $H_\alpha^3$ . To begin, we consider the associated inhomogeneous linear system

$$(2.1) \quad \begin{cases} u_t + u_{xxx} = f, & u(x, 0) = \phi(x), \\ u(0, t) = 0, & u(1, t) = 0, \quad u_x(1, t) = \alpha u_x(0, t), \end{cases}$$

for  $x \in (0, 1)$  and  $t \geq 0$ . Using the notation of the operator  $A_\alpha$  (cf. (1.8)) we rewrite the system (2.1) as an abstract evolution equation

$$\frac{du}{dt} = A_\alpha u + f, \quad u(0) = \phi.$$



Since the operator  $A_\alpha$  generates a continuous semigroup  $S_\alpha(t)$  in the space  $L^2(0, 1)$ , it follows from standard semigroup theory (cf. [16]) that the system (2.1) has a unique solution  $u \in C([0, T]; H_\alpha^3)$  for any  $\phi \in \mathcal{D}(A_\alpha) = H_\alpha^3$  and  $f \in L^1(0, T; H_\alpha^3)$ . The following two propositions present some estimates of these solutions that are needed for establishing the well-posedness of the nonlinear system (1.4).

**Proposition 2.1.** *Let  $T > 0$  be given and  $u(x, t)$  is a solution of the system (2.1). Then,*

$$(2.2) \quad \sup_{[0, T]} \|u(\cdot, t)\|_{L^2(0, 1)} \leq \|\phi\|_{L^2(0, 1)} + \int_0^T \|f(\cdot, \tau)\|_{L^2(0, 1)} d\tau,$$

$$(2.3) \quad \int_0^T u_x^2(0, \tau) d\tau \leq \frac{1}{1-\alpha^2} \left( \|\phi\|_{L^2(0, 1)} + \int_0^T \|f(\cdot, \tau)\|_{L^2(0, 1)} d\tau \right)^2$$

and

$$(2.4) \quad \|u\|_{L^2(-T, T; H^1(0, 1))} \leq c \left( \|\phi\|_{L^2(0, 1)} + \int_0^T \|f(\cdot, \tau)\|_{L^2(0, 1)} d\tau \right)$$

**Proof:** The estimates (2.2)-(2.3) and (2.4) are established using the usual energy estimate method with the multiplier  $2u$  and  $e^{\lambda x}u$ , respectively.

**Proposition 2.2.** *Let  $T > 0$  be given. If  $f \in W^{1,1}(0, T; L^2(0, 1))$  and  $\phi \in H_\alpha^3$ , then the system (2.1) has a unique solution  $u \in C([0, T]; H_\alpha^3)$  for which,*

$$u_t \in L^\infty(0, T; L^2(0, 1)) \cap L^2(0, T; H^1(0, 1)), \quad u_{xt}(0, t) \in L^2(0, T)$$

and

$$(2.5) \quad \|u_t\|_{L^\infty(0, T; L^2(0, 1))} + \|u_{xt}\|_{L^2(0, T; L^2(0, 1))} + \|u_{xt}(0, \cdot)\|_{L^2(0, T)} \\ \leq c (\|\phi\|_{H_\alpha^3} + \|f\|_{W^{1,1}(0, T; L^2(0, 1))}).$$

In addition, if  $f \in L^2(0, T; H^1(0, 1))$ , then  $u \in L^2(0, T; H^1(0, 1))$  and

$$(2.6) \quad \|u\|_{L^2(0, T; H^1(0, 1))} \\ \leq c (\|\phi\|_{H_\alpha^3} + \|f\|_{W^{1,1}(0, T; L^2(0, 1))} + \|f\|_{L^2(0, T; H^1(0, 1))}).$$

**Proof:** It is easy to see that  $v = u_t$  solves

$$\begin{cases} v_t + v_{xxx} = f_t, & v(x, 0) = v_t(x), \\ v(0, t) = 0, & v(1, t) = 0, & v_x(1, t) = \alpha v_x(0, t) \end{cases}$$

where

$$v_t(x) = f(x, 0) - \phi'''(x).$$

Then estimate (2.5) is established in the same way as that to obtain (2.2)-(2.4). If  $f \in L^2(0, T; H^1(0, 1))$ , from the equation  $u_{xxx}(x, t) = f(x, t) - u_t(x, t)$  and the fact

that  $u_t \in L^2(0, T; H^1(0, 1))$  it follows that  $u \in L^2(0, T; H^1(0, 1))$ . Then estimate (2.6) follows from (2.5).

Now we turn to consider that the nonlinear system (1.4) and show that it is locally well-posed using a contraction principle based on the estimates provided above.

**Proposition 2.3.** *For any  $\phi \in H_\alpha^3$ , there exists a  $T = T(\|\phi\|_{H_\alpha^3}) > 0$  such that (1.4) has a unique solution*

$$(2.7) \quad u \in C([0, T]; H_\alpha^3) \cap L^2(0, T; H^1(0, 1)).$$

Moreover, for any  $T' < T$  there exists a neighborhood  $U$  of  $\phi$  in  $H_\alpha^3$  such that the map from the initial data  $\phi$  to the solution  $u$  is Lipschitz continuous from  $U$  to the space class defined by (2.7).

**Proof:** Let

$$Y_T = \{v \in C([0, T]; H_\alpha^3) \cap L^2(0, T; H^1(0, 1)), \\ v_t \in L^\infty(0, T; L^2(0, 1) \cap L^2(0, T; H^1(0, 1)))\}$$

which is a Banach space equipped with the norm

$$\|v\|_{Y_T} := \|v\|_{C([0, T]; H_\alpha^3)} + \|v\|_{L^2(0, T; H^1(0, 1))} + \|v_t\|_{L^2(0, T; H^1(0, 1))} + \|v_t\|_{L^\infty(0, T; L^2(0, 1))}.$$

For given  $\phi \in H_\alpha^3$ ,

$$(2.8) \quad \begin{cases} u_t + u_{xxx} = -vv_x, & v(x, 0) = \phi(x) \\ u(0, t) = 0, \quad u(1, t) = 0, \quad u_x(1, t) = \alpha u_x(0, t), \end{cases}$$

defines a map  $\Gamma$  from the space  $Y_T$  to  $Y_T$ :  $u = \Gamma(v)$  for any  $v \in Y_T$  where  $u$  is the corresponding solution of (2.8). Applying Proposition 2.1-2.2 to (2.8) yields that

$$\begin{aligned} \|\Gamma(v)\|_{Y_T} &\leq c\|\phi\|_{H_\alpha^3} + c\|vv_x\|_{W^{1,1}(0, T; L^2(0, 1))} + c\|vv_x\|_{L^2(0, T; H^1(0, 1))} \\ &\leq c\|\phi\|_{H_\alpha^3} + cT^{1/2}\|v\|_{Y_T}^2. \end{aligned}$$

Choose

$$(2.9) \quad M = 2c\|\phi\|_{H_\alpha^3} \quad \text{and} \quad T = (2cM)^{-2}.$$

Then  $\Gamma$  maps the ball  $S_M = \{v \in Y_T; \|v\|_{Y_T} \leq M\}$  into itself. In addition, a similar argument shows that

$$\|\Gamma(v_1) - \Gamma(v_2)\|_{Y_T} \leq \frac{1}{2}\|v_1 - v_2\|_{Y_T}$$

for any  $v_1, v_2 \in S_M$ . Hence the map  $\Gamma$  is a contraction in the ball  $S_M$  if  $T$  and  $M$  are as chosen in is the needed solution of (1.4). The proof is complete.  $\square$

Next we consider regularity problem of (1.4). Roughly speaking, we show that if  $\phi$  is a smooth function, so is the solution. This would be a consequence of the following proposition to the linear problem

$$(2.10) \quad \begin{cases} u_t + (uv)_x + u_{xxx} = f, & u(x, 0) = \phi(x), \\ u(0, t) = 0, & u(1, t) = 0, & u_x(1, t) = \alpha u_x(0, t). \end{cases}$$

**Proposition 2.4.** *Let  $T > 0$  be given and  $w \equiv w(x, t) \in Y_T$ . Then (2.10) has a unique solution  $u \in Y_T$  for any  $\phi \in H_\alpha^3$  and  $f \in W^{1,1}(0, T; L^2) \cap L^2(0, T; H^1)$ . Moreover*

$$(2.11) \quad \|u\|_{Y_T} \leq c \left( \|\phi\|_{H_\alpha^3} + \|f\|_{W^{1,1}(0, T; L^2(0, 1))} + \|f\|_{L^2(0, T; H^1(0, 1))} \right)$$

where  $c > 0$  depends continuously on  $\|w\|_{Y_T}$ .

**Proof:** For given  $\phi$  and  $f$ ,

$$(2.12) \quad \begin{cases} u_t + u_{xxx} = -(vw)_x + f, & u(x, 0) = \phi \\ u(0, t) = 0, & u(1, t) = 0, & u_x(1, t) = \alpha u_x(0, t) \end{cases}$$

defines a map  $\Gamma$  from  $Y_{T^*}$  to  $Y_{T^*}$ :  $u = \Gamma(v)$  for any  $v \in Y_{T^*}$ , where  $u$  is the corresponding solution of (2.12). Using Proposition 2.1 and Proposition 2.2,

$$\|\Gamma(v)\|_{Y_{T^*}} \leq c \left( \|\phi\|_{H_\alpha^3} + \|f\|_{W^{1,1}(0, T^*; L^2(0, 1))} \right) + c(T^*)^{1/2} \|w\|_{Y_{T^*}} \|v\|_{Y_{T^*}}.$$

Thus if one chooses  $T^* > 0$  such that

$$(2.13) \quad \begin{aligned} c(T^*)^{1/2} \|w\|_{Y_{T^*}} &= 1/2 \quad \text{and} \\ b &= 2c \left( \|\phi\|_{H_\alpha^3} + \|f\|_{W^{1,1}(0, T^*; L^2(0, 1))} + \|f\|_{L^1(0, T^*; H^1(0, 1))} \right), \end{aligned}$$

then for any  $v \in Y_{T^*}$  with  $\|v\|_{Y_{T^*}} \leq b$ ,

$$\|\Gamma(v)\|_{Y_{T^*}} \leq b.$$

Thus  $\Gamma$  maps the ball  $S_b := \{v \in Y_{T^*} : \|v\|_{Y_{T^*}} \leq b\}$  to itself. Similarly, one can also show that  $\Gamma$  is a contraction in the ball  $S_b$ . Its fixed point is the solution of (2.10). Note that  $T^*$  determined by (2.13) only depends on  $w$  and, in particular, not on  $\phi$  and  $f$ . Thus a standard argument can be used to show that  $T^*$  may be extended to any large previously given number and (2.11) holds. The proof is complete.  $\square$

Define a series of differential operators  $Q_k$  for  $k = 0, 1, 2, \dots$

$$(2.14) \quad \begin{cases} Q_0 \phi = \phi \\ Q_{k+1} \phi = -\partial_x^3(Q_k \phi) - \frac{1}{2} \partial_x \left( \sum_{j=0}^k \binom{j}{k} Q_j \phi Q_{k-j} \phi \right) \end{cases}$$

**Proposition 2.5.** *Let  $k \geq 0$  be a given integer. Assume that*

$$Q_j \phi \in H_\alpha^3, \quad j = 0, 1, \dots, k$$

*Then there exists a  $T = T(\|\phi\|_{H_\alpha^3}) > 0$  such that the system (1.4) has a unique solution*

$$\partial_t^j u \in Y_T, \quad j = 0, 1, 2, \dots, k.$$

*As a consequence,*

$$u \in C([0, T]; H^{3(k+1)}(0, 1)) \cap L^2(0, T; H^{3(k+1)-1}(0, 1)).$$

**Proof:** It is Proposition 2.3 if  $k = 0$ . We only provide a proof here for the case  $k = 1$ . The general case follows easily then from induction.

Let  $v = u_t$ . It solves

$$(2.15) \quad \begin{cases} v_t + (uv)_x + v_{xxx} = 0, & v(x, 0) = Q_1 \phi, \\ v(0, t) = 0, & v(1, t) = 0, & v_x(1, t) = \alpha v_x(0, t). \end{cases}$$

It is from Proposition 2.4 that  $u_t = v \in Y_T$ . Then from  $u_{xxx} = -u_t - uv_x$  it follows

$$u \in C([0, T]; H^6(0, 1)) \cap L^2(0, T; H^7(0, 1)).$$

The proof is complete.  $\square$

Now we turn to show globally well-posedness of (1.4) by establishing the following global *a priori* estimates for solutions of (1.4).

**Proposition 2.6.** *Let  $T > 0$  be given and  $u$  be a smooth solution of (1.4). Then*

$$(2.16) \quad \int_0^1 u^2(x, t) dx + (1 - \alpha^2) \int_0^T u_x^2(0, \tau) d\tau \leq \int_0^1 \phi^2(x) dx$$

*and*

$$(2.17) \quad \int_0^T \|u(\cdot, t)\|_{H^1(0, 1)} dt \leq g_1(\|\phi\|_{L^2(0, 1)}, T)$$

*for any  $0 \leq t \leq T$ , where  $g_1(\cdot, \cdot)$  is a nonnegative continuous function.*

**Proof:** This is a nonlinear version of Proposition 2.1 but with a similar proof.

It seems natural that the next step should be finding a global  $H^1$  estimate. However, it turns out being very difficult, if it is not possible, to obtain a global  $H^1$  or  $H^2$  estimate of the solutions of the system (1.4). Instead, we directly obtain the  $H^3$  estimate by estimating the  $L^2$  norm of  $u_t$ . The following two technique lemmas are standard (cf. [1] and [12]).

**Lemma 2.1.** *There exists a constant  $c > 0$  such that for any  $g \in H^1(0, 1)$  satisfying either  $\int_0^1 g(x)dx = 0$  or  $g(0) = g(1) = 0$ , one has*

$$\sup_{x \in (0, 1)} |g(x)| \leq c \|g\|_{L^2(0, 1)}^{1/2} \|g\|_{H^1(0, 1)}^{1/2}.$$

**Lemma 2.2.** *For given positive integer  $m$ , there is a constant  $c > 0$  such that*

$$\|g\|_{H^j(0, 1)} \leq c \|g\|_{H^m(0, 1)}^{j/m} \|g\|_{L^2(0, 1)}^{(m-j)/m}$$

for any  $g \in H^m(0, 1)$  and  $0 \leq j \leq m$ .

**Proposition 2.7.** *Let  $T > 0$  be given and assume that  $u$  is a smooth solution of (1.4). Then*

$$(2.18) \quad \sup_{[0, T]} \|u(\cdot, t)\|_{H_0^3} + \|u\|_{L^2(0, T; H^4(0, 1))} \leq g_2(\|\phi\|_{H_0^3}, T)$$

where  $g_2(\cdot, \cdot)$  is a nonnegative continuous function.

**Proof:** Let  $v = u_t$ . It solves (2.15). Multiply both sides of the equation (2.15) by  $2e^{\lambda x}v$  and integrate over the domain  $(0, 1) \times (0, t)$ . Integration by parts leads to

$$(2.19) \quad \begin{aligned} & \int_0^1 e^{\lambda x} v^2 dx + (1 - e^{\lambda} \alpha^2) \int_0^t v_x^2(0, \tau) d\tau + 3\lambda \int_0^t \int_0^1 e^{\lambda x} v_x^2(x, \tau) dx d\tau \\ & + \int_0^t \int_0^1 e^{\lambda x} v^2(x, \tau) (u_x(x, \tau) - \lambda u(x, \tau) - \lambda^3) dx d\tau \\ & = \int_0^1 e^{\lambda x} \psi^2(x) dx \end{aligned}$$

where  $\psi = Q_1 \phi$ . Taking  $\lambda = 0$  in (2.19) yields

$$(2.20) \quad \int_0^1 v^2(x, t) dx + (1 - \alpha^2) \int_0^t v_x^2(0, \tau) d\tau \leq \int_0^t \int_0^1 v^2 |u_x| dx d\tau + \int_0^1 v^2 dx.$$

Using Lemma 2.1 and Lemma 2.2,

$$\begin{aligned} \int_0^1 v^2 |u_x| dx & \leq \|u_x\|_{L^\infty(0, 1)} \|v\|_{L^2(0, 1)}^2 \\ & \leq c \|u\|_{H^2(0, 1)}^{1/2} \|u\|_{H^1(0, 1)}^{1/2} \|v\|_{L^2(0, 1)}^2 \\ & \leq c \|u\|_{H^4(0, 1)}^{3/8} \|v\|_{L^2(0, 1)}^2, \end{aligned}$$

and similarly,

$$\int_0^1 v^2 |u| dx \leq c \|u\|_{H^4(0, 1)}^{1/8} \|v\|_{L^2(0, 1)}^2.$$

Noticing that  $v = u_t = -uu_x - u_{xx}x$ , one obtains

$$\begin{aligned} \int_0^1 v^2 dx &= \int_0^1 (uu_x + u_{xx}x)^2 dx \\ &\leq \int_0^1 u^2 u_x^2 dx + \int_0^1 u_{xx}^2 x^2 dx + 2 \int_0^1 |u| |u_x| |u_{xx}x| dx \\ &\leq c \|u\|_{H^1(0,1)}^{12/8}. \end{aligned}$$

Consequently,

$$(2.21) \quad \int_0^1 v^2 (|u_x| + |u|) dx \leq c \|u\|_{H^1(0,1)}^{15/8}.$$

In addition, since  $v_x = u_{xt} = -u_x^2 - uu_{xx} - u_{xxx}x$ ,

$$\begin{aligned} \int_0^1 (v_x^2) dx &= \int_0^1 (u_x^2 + uu_{xx} + u_{xxx}x)^2 dx \\ &= \int_0^1 (u_{xxx}x)^2 dx + \int_0^1 u^2 (u_{xx})^2 dx + \int_0^1 u_x^4 dx + 2 \int_0^1 u_{xxx}x u_{xx} u dx \\ &\quad + 2 \int_0^1 uu_x^2 u_{xx} dx + 2 \int_0^1 u_x^2 u_{xxx}x dx. \end{aligned}$$

Using Lemma 2.1 and Lemma 2.2,

$$\begin{aligned} \left| \int_0^1 u_{xxx}x uu_{xx} dx \right| &\leq \|u_{xx}\|_{L^\infty(0,1)} \|u\|_{L^2(0,1)} \|u\|_{H^1(0,1)} \\ &\leq c \|u\|_{H^3(0,1)}^{1/2} \|u\|_{H^2(0,1)}^{1/2} \|u\|_{H^1(0,1)} \\ &\leq c \|u\|_{H^1(0,1)}^{13/8}. \end{aligned}$$

Similarly,

$$\left| \int_0^1 uu_x^2 u_{xx} dx \right| \leq c \|u\|_{H^1(0,1)}^{7/4}, \quad \left| \int_0^1 u_x^2 u_{xxx}x dx \right| \leq c \|u\|_{H^1(0,1)}^{15/8},$$

and

$$\int_0^1 u^2 u_{xx}^2 dx \leq c \|u\|_{H^1(0,1)}^{5/4}, \quad \int_0^1 u_x^4 dx \leq c \|u\|_{H^1(0,1)}^{5/4}.$$

Thus

$$(2.22) \quad \int_0^1 v_x^2 dx \geq \int_0^1 u_{xxx}^2 x^2 dx - c \|u\|_{H^1(0,1)}^{15/8}.$$

We obtain

$$(2.23) \quad \int_0^1 v^2 dx + 3\lambda \int_0^t \int_0^1 u_{xxxx}^2 dx d\tau \leq \lambda^3 c\lambda \int_0^t \int_0^1 v^2 dx d\tau + c(\lambda) \int_0^t \left( \int_0^1 u_{xxxx}^2 dx \right)^{15/8} d\tau + \|\phi\|_{H_\alpha^3}^2$$

for any  $0 \leq t \leq T$ , which implies (2.18). The proof is complete.  $\square$

**Remark 2.1.** As a consequence of continuous dependence of solution on its initial data, (2.21) is not only true for smooth solutions but also true for any solution in the class of  $C([0, T]; H^s(S))$  for  $s \geq 3$ .

The following theorem now follows from a standard argument.

**Theorem 2.1.** Let  $T > 0$  be given. For any  $\phi \in H_\alpha^3$ , there exists a unique solution  $u \in C([0, T]; H_\alpha^3)$  to (1.4). Moreover, the map from the initial data  $\phi$  to the solution  $u$  is Lipschitz continuous from  $H_\alpha^3$  to the space  $C([0, T]; H_\alpha^3) \cap L^2(0, T; H^1(0, 1))$ .

**Remark 2.2.** In fact, as in [19], we can show the map  $K$  is analytic from the space  $H_\alpha^3$  to the space  $C([0, T]; H_\alpha^3) \cap L^2(0, T; H^1(0, 1))$ .

As a corollary of the above theorem and the Proposition 2.1, we have the following regularity result.

**Corollary 2.1.** Let  $T > 0$  be given and  $k$  be a positive integer. If  $Q_j \phi \in H_\alpha^3$ ,  $j = 1, 2, \dots, k$ , then (1.4) has a unique solution  $\partial_t^j u \in Y_T$  for  $j = 0, 1, \dots, k$ , and consequently,

$$u \in C([0, T]; H^{3(k+1)}(0, 1)) \cap L^2(0, T; H^{3(k+1)+1}(0, 1)).$$

### 3. Spectral Properties of the Operator $A_\alpha$

The operator  $A_\alpha$  defined in section 1 generates a strongly continuous semigroup  $S_\alpha(t)$  on  $L^2 \equiv L^2[0, 1]$  for  $t \geq 0$ . For the work to be done in this paper we require very detailed information about the eigenvalues of  $A_\alpha$  and its adjoint  $A_\alpha^*$ , which agrees with (1.8) except for removal of the “-” sign there and reversal of the roles played by 0 and 1 in the boundary conditions. Since  $A_\alpha$  for  $|\alpha| < 1$  is a dissipative differential operator, it has compact resolvent and therefore its spectra consists of only countable many eigenvalues  $\lambda_k$  with  $\operatorname{Re} \lambda_k \leq 0$  for  $k = 0, \pm 1, \pm 2, \dots$